

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2002, 13:13:05 ; Search time 5580.66 Seconds  
(without alignments)  
17646.709 Million cell updates/sec

Title: US-09-842-930A-1

Perfect score: 4706  
Sequence: 1 tcttaccagaagctcactcac.....aagcaaaaaaaaaaaaaa 4706

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htgo\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	2875.2	61.1	8266	9	HSA295695	AJ295695 Homo sapi
2	2823.6	40.0	4575	9	AK024503	AK024503 Homo sapi
3	2059.6	43.8	3642	9	HSMB01377	HSMB01377 Homo sapi
4	1936.8	41.2	3260	9	AF160476	AF160476 Homo sapi
5	1599.4	36.5	3625	6	AX149461	AX149461 Sequence
6	1244.8	26.0	2483	6	AX149459	AX149459 Sequence
7	1243.2	26.4	2512	6	AX149455	AX149455 Sequence
8	1196	25.4	2011	6	AX149447	AX149447 Sequence
9	1016.4	21.6	1804	6	AX149449	AX149449 Sequence
10	698.8	14.8	6777	9	D87433	D87433 Homo sapien
11	697.2	14.8	7870	9	HSB275213	AJ275213 Homo sapi
12	637.6	13.5	1624	6	AX149457	AX149457 Sequence
13	517.4	11.0	1577	6	AX149463	AX149463 Sequence
14	154	3.3	218073	2	AC025501	AC025501 Mus muscu
15	146.8	3.1	157216	2	AC022277	AC022277 Homo sapi
16	146.8	3.1	180064	2	AC025265	AC025265 Homo sapi
17	146.8	3.1	180525	2	AC063946	AC063946 Homo sapi
18	146.8	3.1	207187	2	AC012555	AC012555 Homo sapi
19	113	2.4	218073	2	AC025501	AC025501 Mus muscu
20	112.2	2.4	143387	2	AC099529	AC099529 Felis cat
21	111.8	2.4	157452	2	AC010190	AC010190 Homo sapi
22	104.4	2.2	237252	2	AC099543	AC099543 Homo sapi
23	103.8	2.2	157216	2	AC022277	AC022277 Homo sapi
24	97.8	2.1	237252	2	AC099543	AC099543 Homo sapi
25	91.6	1.9	1249	10	BC022136	BC022136 Mus muscu
26	83.4	1.8	340	11	G24845	G24845 human SRS E
27	83.4	1.8	370	11	G29550	G29550 human SRS E
28	79	1.7	1406	4	RABPS4EA	M6381 Oryctolagus
29	78.6	1.7	216291	2	AC105934	AC105934 Homo sapi
30	77.4	1.6	1605	10	MM083903	U83903 Mus muscu
31	77.4	1.6	1616	10	BC021155	BC021155 Mus muscu
32	70.8	1.5	834	9	HSB419936	AJ19936 Homo sapi
33	70.8	1.5	834	9	HSB421518	AJ21518 Homo sapi
34	70.8	1.5	1414	6	AR063905	AR063905 Sequence
35	70.8	1.5	1414	6	ARI44434	ARI44434 Sequence
36	69.6	1.5	769	9	HUM2D93D04	AF086484 Homo sapi
37	67.6	1.4	1414	6	AX336169	AX336169 Sequence
38	67.6	1.4	1414	6	HUMTSG6A	M31165 Human tumor
39	63	1.3	1411	6	AX202114	AX202114 Sequence
40	61.6	1.3	164286	9	AC020899	AC020899 Homo sapi
41	61.6	1.3	167872	2	AC007554	AC007554 Homo sapi
42	61.2	1.3	1478	4	ECCTRL1G	X78077 E. caballus
43	60.2	1.3	3633	4	AB050947	AB050947 Felis cat
44	60	1.3	1400	6	AR036272	AR036272 Sequence
45	60	1.3	1400	6	AR087156	AR087156 Sequence

#### ALIGNMENTS

RESULT 1  
HSA295695  
LOCUS HSA295695 8266 bp mRNA linear PRI 04-JAN-2002  
DEFINITION Homo sapiens mRNA for stabilin-2 (STAB2 gene).  
ACCESSION AJ295695  
VERSION AJ295695.1 GI:18073427  
KEYWORDS STAB2 gene; stabilin-2.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiina; Homo.  
AUTHORS

TITLE  
Guillot, P., Johannson, S., Birk, R., Hakily, N., Franke, P., Kodejta, V.,  
Kannicht, C., Orfanos, C.E., Johannson, S. and Goerd, S.  
Stabilin-1 and stabilin-2 constitute a novel family of fasciclin  
domain-containing adhesion molecules associated with endothelial-  
macrophage differentiation and angiogenic processes

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (27-JUL-2000) Gratchev A., Department of Dermatology,  
Unpublished  
2 (bases 1 to 8266)  
Gratchev, A.  
Direct Submission



Db	4731	CCGCTGTTTGGAGAACACATGTTGGCTGTGACAAAGAAATGGGGAGTGGACACAGACAGGACC	47930
Qy	1197	caaccagggccgclgtctaactgcttgcgcgaagatcacatggaatggaaagtgtgcgtc	1256
Db	4791	CAACCAAGCGCTGCTGTAACTGTTGGCACATACACTGAGATGGAAAGTCTGGACACT	4850
Qy	1257	latcaatgcttgccttaacgaacaatgagcgctgcagatccatttgccttgcctaataac	1316
Db	4851	CATCAATGCTGTGCTTAACTAAATAATGGCGGCTGATGTAATTTGCCATCTCCACACAC	4910
Qy	1317	tgaaagaatlcgaaggaatgatctccggaagccagatcacagggatgaaatgcgtc	1376
Db	4911	TGGGCAAGTAAAGAAAGCACTTGTACTTGCAAGCCAAACTACATTTGAGATGGCATTTACCTG	4970
Qy	1377	ccggaggaagcatcctacgaggaagcttcccaagaaccttcgaagctccagtaacttcca	1436
Db	4971	CCGGCGAGCATTTATCAGAGCTTCCCAAGAAACCGAATACTCCAGTATTTCTTCCA	5030
Qy	1437	gttcgaggaagcaatgctgtlccggaagcttgcctggaaccttgcccctacccgtgtccgc	1496
Db	5031	GTTTCAGAGACATTTGCTGTGAAGATCGTGCGGCCAGGCCCTTCACGTGTTTGGACCC	5090
Qy	1497	ttttcctaagcttcaatcatatgagcccggttaaagaattggatcagagggccatc	1556
Db	5091	TTTTATCTCGACGCTTTTGATGAGGAGAGCTGGGGTTAAAGACTGGGCAAAATACGGTTTAA	5150
Qy	1557	gtcccaagcttcttgcgtatacagtggtggcgctgcgcagcagctgcgcgtttggacaactaa	1616
Db	5151	GGCCGACGTTCTTGGTACCATGTGGTGGCTGGCCGACCAAGCTGCTTGTGGAAACCTGAA	5210
Qy	1617	agtgacacacaagtgcgaagacccttccaaagagagaccagtttccatctgtctctcaga	1676
Db	5211	ATTGATGTCAAATGCTACTTCCCTCCCAAGAGAGCCAAATGACATCTCCGTCGTCCAGAG	5270
Qy	1677	caactgttcatataacaatgagaggaaggttccctgtccagtgcatacatcagaacaatgg	1736
Db	5271	CACGGTGTATTAATAATAATAGGCTTAAGATCATATCCATGATATCATACAGTACTAAATGG	5330
Qy	1737	cgatcacagcttatagacaagctgtgtcttccaaaaactgtgtatcacccccaaaga	1796
Db	5331	GATTGTTCAATATCATATGACAAATTTGCTATCTCCCAAAATTTTGCTTATCACTCCCAANGA	5390
Qy	1797	tgccctgggcagaggtctctgcgaatcttactacagtgcgcgaagaacacagatalacaa	1856
Db	5391	CAACTCTGGAAGAAATTTCTGCAAAATCTTACGACTTTGGCAACAAACATGGCTACATCAA	5450
Qy	1857	atcagaagaattgatatacagaagcttcagagcttgcgttcagatcatcactgactcatalccaac	1916
Db	5451	ATTTAGCAACTTATATACAGGACCTCAGGTTTGGTGAATGCATACCGATCCCATCCATCCAC	5510
Qy	1917	cccaagtcacgtcttcttgcgcctcgggaagaagccctggagaagctctgccccagaagca	1976
Db	5511	CCCAGTACTCTCTTCTGGGCCACCGACCAAGGCCCTTCCATGCGCTACTGCTGAACAACA	5570
Qy	1977	ggaccttccgttcaatcaagaacaacaagaagaagcttgaagcttcttccctgaagttccag	2036
Db	5571	GGAATTCCTGTTCAACCAAGACAAACAAGACAAGCTGAAGAGATTTTGAAGTTTACAT	5630
Qy	2037	gattccgaagctccaagagctttagctcagaacctcccaagcttgccttcctcgaagacct	2096
Db	5631	GATACGAGATGCCCAAGTTTGTACTGTTGATCTTCCCATCATCCATCGCTCGAAGACCT	5690
Qy	2097	gcaagagctcgaagcttaagatgtagaggtgtgaactctggaatgagacatccgtagagctcttct	2156
Db	5691	GCAAGGTTTCAGAACTGAGTGTGAATGTTGAGACTGGCAAGGACATCGGTGTGACTCTTTCT	5750
Qy	2157	aaacgaacaatgttgagaattcacaacacgggaagctctgtttgacgtgggtgtgcgcta	2216
Db	5751	GAAATGGCAAACTCGCAGAAATTTGTGAGGGGAGGAGCTCTTGTTTGAACCTGGGTGTGGCTA	5810
Qy	2217	tgacattgactgtcctacatgaatccctacacctgaagtggcgatgtagaacatttactac	2276
Db	5811	CGGATGATGATGCTCTCTGATTGATCCACCTCTGGGGGGCGGCTGTGACACTTATAC	5870

Qy	2277	ctgcatatccgggagagagcggaagtgtgcacttccctcccaaatgcccacgtgaag	2330
Db	5871	TTTCGATGCTCCGGGGAGTGTGGAGCTGTGTCAATACTCCACAGTCCCAAGCTGAG	5930
Qy	2337	caagccaagagcgctaaagaagaagtgtatctacaaccgttaccttcaagaaagcgt	2396
Db	5931	TAAACCAAGGGTGTGAAGCAGAAGTGTCTTCAACACC---TGGCTTTCAAGACAGAACT	5987
Qy	2397	ggaaagctgcacgaacactgtgcaacgttgatccaaccccccaagtgctgtgcaagtta	2456
Db	5988	GGAAAGCTGGCCGGAGAGGGTGCAGCTGTGTATACAGATATCCCAAGTCTGTCAAGGGCTA	6047
Qy	2457	cttcaatgcagacactgtgcaagccctggagagaccagataacacggtttaacaaacggg	2516
Db	6048	CTTGCGGCGAGACTGTGCAAGGCTGCGCTTGAAGGACCAATGCCCCGTATTAAACCGGGG	6107
Qy	2517	catgtgcgcgacatctgtacaacccaatggagacgtgctatgtgcacaacacgggtctcaag	2576
Db	6108	TGTGTGCTTGATGATCAGTACGTACGAGCCACCGAGAGAGTGTAAATGCAACACGGCTTCATATG	6167
Qy	2577	gaacagcttggaagctctgtctgagaaatggagatattgggacctacgtctgaagccccgaagt	2636
Db	6168	GACGGCTGTGACATGTGCTGGCCGGGGAGATTTCGGCTCAATTTGTCGCTCTGTGGCTG	6227
Qy	2637	ctccgaagcatgtgacagatgtatataagggatatacaagctccggagatgcctctgtgaac	2696
Db	6228	CTCAGACCAAGGAGACATGGCATGTATGTGATCAACGGGGCTCCGGGCAAGTCCCTCTGTGAAC	6287
Qy	2697	agggtgtgaacagccgctctgtgtgaacactcccaacagctgatatcgagtggtgacacctgc	2756
Db	6288	GGGGTGTGACAGGGCCCTCGTGTACACTAGGCACATTTTGGCTGTGAGTGTACGCTTCC	6347
Qy	2757	tgtctctgcagcaagcaccctgttagggaaagaacaacggtgtgtgtgaactgtgaactcaga	2816
Db	6348	TTGTCTTGTCTATATGCCACTGTATAGAGACACACAGGTGATGATTAACCTGTGAATTATGA	6407
Qy	2817	aggtgtacggatatacatgtgcaacagctgtggaactctgtgcaacaagaacaacaggggactgtgc	2876
Db	6408	AGGTGACGGATACATGCACACTTGTGTGATTTGTCAAACAGGACAAAGGGGGCTGTGC	6467
Qy	2877	gaagtgctgtcaaatgtgtctcccaagaagaagcaccgaatctctctgcagctgtgcaagaagcta	2936
Db	6468	AAAGGTGTGCCAGATGTCTCCACAGAGGGCCAGAAAGTTCCTGTGCAGCTGCCAAGAGGGATTA	6527
Qy	2937	caaggggagatgtgactacaagctgtgcaataagaataaacctctgtgcaacaggtgtgtcaacgggg	2996
Db	6528	CAAGGGGACGGGACACAGCTGCACAGATATACCCCTGTGCAGACGGCTTAAAGSGAGG	6587
Qy	2997	atgcacatgacacgcacactgtgcagatgtgacgggcccgaagcaagcataagtgtgaatgttaa	3056
Db	6588	GTTGTCAAGACACGCCACACTGTATAGATGTGACAGCCCGGGCAAGATGTGTGATGTATA	6647
Qy	3057	aagtaacatagtgtgggaagcagagtgtaagctgtgaagcactgtgcgcgtccagacgtgtg	3116
Db	6648	AAGTCACTATGTGTGGAGATGTGGCTGTGAAGCTGTAGCCGGAGACAGCTGCCATTTAGACCCCTG	6707
Qy	3117	cttacaagacaacgcgacagtgaccacccaatgtgacacgctgtgcaagactctacttaccaga	3176
Db	6708	CTTACAGGACAAATGGGCGAGTGGCCATTCACAGCCAAATGTGTGTGCAGCTTCCAGGA	6767
Qy	3177	caagacccgtatggagtattccatctaacgctccccaacttgggccaagtatacaaatctgaacttga	3236
Db	6768	TATCACATGTTGGGGTGTCCATCTATACGGTCCCACTGTGGGCCAAGTATTAAGCTGACCTTTGA	6827
Qy	3237	caaaagcaaaagagcctgtgtgcaaaagaagctgtgcagcaatgacaacacttacaacagccttc	3296
Db	6828	CAAAAGCCAGAGAGCTGTGTGCCAAGAAAGCTGTGGACATGTGCACACTTAACCAACAGAGCTGTC	6887
Qy	3297	ctatgtcccaagaagccaagatatacactgtgtctgcgcagctgtgctgtgaagatgtggcggt	3356
Db	6888	CTATGTCCCAAGAGCCAAAGTACCACTGTGTCTACACAGGCTGTGGATGTGAAGACCGGGCGGT	6947







QY 2163 acaaatgtgcagatcatcatcaccggagacatctgttgcagtcggtgtgtgactatgacat 2222  
DB 2099 CCAACCTGCAAAATGTCACGCGAGCTCTGTTTGACGTGGGTGTGCTACGCGCAT 2158  
QY 2223 tgcagcttaccatcgaatccctaccatggtgtgcgaatgtgaacttcttaccatctga 2282  
DB 2159 TGACAGTCTGCGATGATGATCCACCTGGGGGCGCTGTGACACCTTTACTACTTTGCA 2218  
QY 2283 tattccggggaggtgcggaggtgtgcatcttccatcccaatgccactgaagcaagc 2342  
DB 2219 TCCCTCGGGGGAGTGTGGGAGCTGTGTCAATCTCCACTGTGCCAAGTGGAGTAAC 2278  
QY 2343 aaaaagcgctgaagaagaagtgtatctacaaacccgttacccttccaggaggaagcgtgaag 2402  
DB 2279 AAAGCGTGTGAAGCAGAAATGTCTCTACAACC---TGCCCTTCAAGAGGAACTGTGAAG 2335  
QY 2403 ctgcagaacacctgtgcacccgtgtgtatccaaaccccccaagtgctgcagatgttactcat 2462  
DB 2336 CTGCGGGAGCGGTGACGCTGGTATACGATCCCAAGTCTGTGCAAGGGCTACTTTCG 2395  
QY 2463 gccagactgtcaggccctgcggaggacagatacacccgtgtaaacacgggcagtgtc 2522  
DB 2396 GCGAGACTGTACGCGCTGCCCTGGAGAGACCATGCCCCGTGTAATAACCGGGGTGTG 2455  
QY 2523 ccgcagactgtacacacccatgaggacagtgtcctatgcacacccggtctcaacggagac 2582  
DB 2456 CCTTGATAGTACTCGGACCGGAGAGTGTAAATGCAACACCGGCTTCAATGGAGCGC 2515  
QY 2583 ctgcagacctgtgtgcagatggagagattgggacctgtcaagccccgaagctgtccga 2642  
DB 2516 GTGTGATGTCTGTGCGGGAGATTTGGCTGTGATGTCTGTGCTGTGGCTGTGCTACA 2575  
QY 2643 gcatgagacagtgtgataaggagatcacagctccggaggagtgctctgtgaacaagtg 2702  
DB 2576 CCACGGACAGTGCAGATGATGATCATCGGGCTCCGGGAGTGTGTGGAAGCGGG 2635  
QY 2703 gacagccgctctgtgtgacactcccaagctgtatctgcagtggtgcacacctgtgtctc 2762  
DB 2636 GACAGGCCCTCGTGTGACACTGACACTTGTGCTGTGCTGTGCTGTGCTGTGCT 2695  
QY 2763 cgtgcagccacctgtgagagagaacacgtgtgtgtgttaactgtgaactgaaggtga 2822  
DB 2696 TCCCTATGCGACCTGTGAAGGAGAACACGTTGTGATTAACCTTGATTAAGAAGTGA 2755  
QY 2823 ccggaatcatgtcacagctgtgtgactctgtcaaaacgaacaacggggtgtgtgcaggt 2882  
DB 2756 CCGAATCATATGACAGTGTGTGATTTCGCAACGACGACGAGGGGCTGTCCAAAGT 2815  
QY 2883 cgtcaagtgtccacagaagaaccccaagctctctgtcagctgtcagaagaagctacaagg 2942  
DB 2816 GCGCAGATGCTCCAGAAAGGACGCAAGGCTCTCTGACCTGCAAGAGGATACAAAG 2875  
QY 2943 ggaatgctcagctgtgatatagataagccctgtgagagcggtgtgaaggaggatgcca 3002  
DB 2876 GGAAGGGACAGCTGCACAGAGATGAACCCCTGTGCAGAGGCTTTAAAGGAGGGTGA 2935  
QY 3003 tgaagccagccactgtcagagatgacggccagcagcaagcaatggtgtgaatgttaaaagtca 3062  
DB 2936 CAGAGCAGCGCACTGTAATATATACAGCCCGGCAAGCAAGTGTGAGTGTAAAGTCA 2995  
QY 3063 ctatgtcggggagcagagtgagactgtgagctgtgagcagctgtgcagcgttgtctaca 3122  
DB 2996 CTATGTGCGAGATGGCTTAACCTGTAGCGGAGCTGTGCCATTGACCGGTGCTTACA 3055  
QY 3123 ggaacaagcagatgtgcacccagatgtgcagctgtgtgagagccttacttccagacagc 3182  
DB 3056 GGAATATGGGAGTGTGCACAGACGCAAAATGTGTGACCTTCATTCAGAGTACCAAC 3115  
QY 3183 cgttagagatcatctacgtctcccaactggccagatlaaactacatatcttacaagc 3242  
DB 3116 TGTGTGGGTGTTCATCTACGCTCCCACTGGGCGAGTAAAGTCACTTTTACAAAGC 3175  
QY 3243 caaagaagcctgtgtccaaagaagctgtgcacatagccactacaacacagctctccta 3302

DB 3176 CAGAGAGGCTGTGCCAAGAAAGCTGCGACCATGGCAACCTACACACGCTCTCATGC 3235  
QY 3303 ccaagaagccagatctacactgtgtcgtgcggctgtgcgtggagagatgggggggtgtccta 3362  
DB 3236 CCAGAAAGGCCAAGTACACCTGTGTGCTACAGAGCTGTGGAGACCGGGGCTGTGCTTA 3295  
QY 3363 ccgaactagctatgtcctcccaagaagtgtgtgtcaacgtgtgtgtgagatcgtactacgg 3422  
DB 3296 CCCACAGCCTTCCGCTCCCAACACTGTGCTGTGTGTGTGTGGATGTAGTACATAGG 3355  
QY 3423 atccaggcccaacaaagatgtaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3482  
DB 3356 AACTGACCCCAACAAAGAGTGAATGTGGATGTCTCTCTATCGGATGAAGAATGTGA 3415  
QY 3483 ctgcagcttcaaggaagctgt 3542  
DB 3416 CTGACCTGCAAGGTGGGTATGTGTGGAGATGGCTTCTATGCAAGTGGAACTGCTGCA 3475  
QY 3543 ggtccatcatgtccttccctctgcacaaactcctctgacagagatgtgtgtgtgtgtgt 3602  
DB 3476 GGTCTGATGTCTTCCCTCCCACTGCAAACTTCTGTGAGGAAAGTGTGCTGCTATTCGA 3535  
QY 3603 gagctcagcccgaggaagcaatlttgtaaacactgtgactgtgtgtgtgtgtgtgtgtgt 3662  
DB 3536 CAGCTGAGCTGAGGCGGTGATTTCTAGAACACCTGACTGACTGTGCTCATCGGGAC 3595  
QY 3663 cctgt 3722  
DB 3596 CCTCTTGTGTGCAACAGATGTGGCTGTGGGAGATGTGAGTGTGTGTGTGTGTGTGTGT 3655  
QY 3723 tgaagcacctcatcatatgtcaagctcctcttcaaaatgtcaatgtgtgtgtgtgtgt 3782  
DB 3656 CGAGCACCACTTCCCAATGTGACATGTTTCTTCAATGTGACTGTGCAATGGCACAC 3715  
QY 3783 tctgagactatgtctgt 3841  
DB 3716 CCGCAAAAGAGGCTGTGGAGCAAGGTGCTCATCATGTGCCAGCAAGCACTCCAAAC 3775  
QY 3842 --aaagaccaggt 3899  
DB 3776 GACGGAGACAGGTTTGT 3835  
QY 3900 tgaatcctcatatlatcttctgaacctgtgagctctcccaacggcagcaagcgtgtgc 3959  
DB 3836 TGGAGATTTATGATGATTTCCAGGCTTTAAAGCAACCCCTGCGCCGTGACCTTGAC 3895  
QY 3960 ccaactgtgcctgt 4019  
DB 3896 CCACACTGTGCTGTGGAGCAAGGATCTTCTTGTGCAATCATCTGTGTGTGTGTGTGTGT 3955  
QY 4020 tctgagacttacttacttccggtgaagcagcgaacacgtgtgtgtgtgtgtgtgtgtgt 4079  
DB 3956 CTGTGCTGT 4015  
QY 4080 -tcaagaagagacatgtatgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4138  
DB 4016 GTGCGAAGAGACATTAATGTTCACAGCTCTGTGCAAGCAGGACCTGAGAATATCTCGAA 4075  
QY 4139 cccctctgtatgtgagacttaagcgcggcaccaccccaaggtctcctctgtgacaccttaacga 4198  
DB 4076 CCCCTTGTATGAGAGCAACACTGACCTCCCGAGAACCTTCTTCAAGACCCCTTCGGA 4135  
QY 4199 cccctgtgagagacagatctgtgagagacagcctctgtgtgtgtgtgtgtgtgtgtgtgt 4255  
DB 4136 CTCTGAAGAACGGAGCTTGTGAGGCAATGACCCCTTGTGAGGACACTGTGAGGCGCTTGAG 4195  
QY 4256 tgaagaagcagcaagcaacaacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4315  
DB 4196 GGAATGAGCCATCATCTACATGTGCACTGTGGCACTATCAATGTGTAATCTGACGAC--- 4252  
QY 4316 catgtatcagttgtttaaagaatgtacaactctcataaagccagctactactcctctct 4375





[illegible]

Db	911	GTGTGGAGCTGTCTCAATCTCCACCTGCCCAAGCTGAGTAAACCAAGGCTGGAA	970
Qy	2355	gaagaaatgtratacacaaccgttaacttcaaggagaagctggaagcct	2414
Db	971	GCAGAAATGCTCTTCAAC---TGCCCTTCAAGAGAACCTGGAAAGCTCCGGAGCG	1027
Qy	2415	gtgcacggtgtgataccaaccccaagtgctgcatgtacttcaatgccaagtca	2474
Db	1028	GTGAGAGCTGTGATACAGATCCCAAGTGTGCAAGGGCTCTCGGAGAGCTGTCA	1087
Qy	2475	ggcctgcccgggagaccataaccgtttaacaacgggacatgycgcgactgta	2554
Db	1088	GGCCTGCCCTGGAGACCAATGCCGTGTAATACCGGGGTCTGCTTGTACAGTA	1147
Qy	2535	caacacatggagacatgacctatgccacgggttcaacgggacagcctgcgacctg	2594
Db	1148	CTCGGCCACCGAGAGTGTAAATCAACCGGCTTCAATGGACGGCGTGTGAGTGTG	1207
Qy	2595	ctggcatggagatcttgacctgactlcaagcccgacgctgctcagacatggacagtg	2654
Db	1208	CTGGCCGGGGAGATTTCGGGCTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1267
Qy	2655	tgatgagggatcaacagctccggggagtgctctgtgaaacaggtggaagcgccttc	2714
Db	1268	CGATGATGGCATCACGGGCTCCGGGAGTGCCTGTGAAACGGGGTGGACAGGCCCTC	1327
Qy	2715	gtgtgacatcccaacgctgactgacgtgacgtgtgacacccctgtgctccgctgac	2774
Db	1328	GTGTGACACTTACGACAGTTTGGCTGCAAGTGTACGCTCTCTGTTCTGCTCATG	1387
Qy	2775	ctgtacggagaacaacacggtgtgtgtaacttgaactaagaagtgagcggatcacatg	2834
Db	1388	CTGTAAAGAGAAACACAGTGTGAGTAACTGATTAACAGTGAACGGAGACACATG	1447
Qy	2835	caagctgtgagacttctgcaacagaacaaggggcgtgtgcaaggtgcgtaagtgctc	2894
Db	1448	CACAGTTGTGATTTCTGCAACAGAACGGGGGCTGTCAAAAGGGGCAAGTGTCTC	1507
Qy	2895	ccaagaagcacaacaaactctctcctgcaagctgcaagaaggtcacaaaggggata	2954
Db	1508	CCAAGAGGCGCAGAAAGTCTCTGAGTGTGCAAGAGGATACAAAGGGGAGCGGCA	1567
Qy	2955	ctgcataagataagacccctgtgcaagcgtgtgcaaggggagtgcaatgacagccac	3014
Db	1568	CTGCACAGAGATACACCTCTGCAAGAGGCTTAAACGGAGGCTGTCCGAGCACGC	1627
Qy	3015	ctgcagatagcggcccaagcagaagataagtgtaagtgtaaagtcaatgtgcgggga	3074
Db	1628	CTGTAAAGTGAAGGCGCGGAGCAAGCAAGTGTGATGTAAAGTCACTATGTGGAGA	1687
Qy	3075	cgaagtgaactgtagccttgcaagcgtgtgcgctgcaacggtgttcaagaagcaagca	3134
Db	1688	TGGGCTGAACGTGAGCGCGGAGCAAGCTCCCATTTGACCGCTTACAGGCAATGG	1747
Qy	3135	gtgcacaccagaatgcagcgtgtgcaagccttacttccaggaacagcagctagaatatt	3194
Db	1748	GTGCATGCAAGCAGCAAAATGTGTGACCTTCACTTCCAGATACCACTGTGTGGGT	1807
Qy	3195	ccatctacgctcccaactgggacaggtataaacttgaacaaagaaagagcctg	3254
Db	1808	CCATCTAGGCTCCCACTGGGCGCAATTAAGCTGACCTTTGACAAAGCAAGAGGCTG	1867
Qy	3255	tgcacaagaactgcagacatagccaactacaacagcttccclatgccaagaagccaa	3314
Db	1868	TGCCAAGCAACTGCGACATAGCAACTTACACAGCTCTCCATATGCCAAGAGGCA	1927
Qy	3315	gtatcaactgtgtcgcgcgagctgtgagagtggtgcgggtgtgcttaaccgaacta	3374
Db	1928	GTACACACTGTGTGAGAGGCTGTGAGACCGGCGGCTGTGCTTACCCACAGCTT	1987
Qy	3375	tgccttcagaagtggtgtgaaacgttgtggatcgtaactacgattccagcccaa	3434

---

Db	1988	CGCCTCCAGACTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2047
Qy	3435	caagagtgaatgtggaatgtctctgttaccggaatgaagatgtgaactgcacgtgcaa	3494
Db	2048	CAAGAGTGAATGTGGGATGCTTCTGTATGTGATGAAGATGTGAATGTGACCTGCA	2107
Qy	3495	ggcagctatgtggaatgtgctctctgtgcaatgtggaactgtgtgaatgtccatgtc	3554
Db	2108	GTGTGGGTATGTGGAATGTGCTTCTCATGTGAGGGAACCTGTGAGGCTGTGATGT	2167
Qy	3555	cttccctcgtcaacaacttccagagaagtgctgtgttcttccagaagcagccg	3614
Db	2168	CTTCCCTCACTCAAACTTCTGAGGAGTGTGTGCTGTGCTTCTTCAAGAGCTGAG	2227
Qy	3615	aggaacgcaatlttgaacaacactgacatgacatgcaatgcaatgcaatgcaatgca	3674
Db	2228	AGGCCGTGATTTTGTAGAACCTGTGACCTGTGATCCAGCGGCGGAGCTGTGTG	2287
Qy	3675	acagaacagtgtggtctacccggaataaagacgtgtgtgcccgggaatltggaacac	3734
Db	2288	ACAGAACAGTGGGCTGGGGAGTGAAGACCTTGTGGGGGAGCATGAGACACACT	2347
Qy	3735	caactatgcaagcttcccttcaatgacatgcaatgcaatgcaatgcaatgcaatgca	3794
Db	2348	CGCCAAATGTGACATGTTTTTCTACATGACCTTGTCAATGTGACACACCTGCA	2407
Qy	3795	gttggagaagcgaactgtcattacatcttcaagcagaacagctcacc---aagaacag	3851
Db	2408	GGTGGAAGCAAGCTGTGATCACTGTGACAGCAGACCTTCAACAGAGAGAGCAG	2467
Qy	3852	gttgtgtatggaatgcatctctgcaagtggaatatacgcgcaatggaatgccca	3911
Db	2468	GTGTGTGATGGAAGAGCATTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG	2527
Qy	3912	tattatltcgaacacttgaagctcctcccaagcagaagcagctgtcccaactgtgct	3971
Db	2528	TGTCTATTTCAGGCTTTAAAGCACCCCTGCCCCGTGACCTTGAACCACTGGCTT	2587
Qy	3972	ggggaagatlatctgtgcgtgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	4031
Db	2588	GGGAGCAGAGATCTTCTTGTGATCATCTGTGATCTGTGATCTGTGATCTGTGAT	2647
Qy	4032	cttacttccgcttaagcagaacacacacacacacacacacacacacacacacacac	4090
Db	2648	CTTCTACTTGTGATTAACCGGAGAACATGGCTTCCAGATTTTGTGAGTGTGAGGA	2707
Qy	4091	catgtatgtctgtgttctgcaagcagcagcagaatatacgaacacacacacacac	4150
Db	2708	CATTATTTGTGACCTTGTGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	2767
Qy	4151	gaactcagcgcgcgcaaccccaagatgctcctgtgacccctcaagaacacacacac	4210
Db	2768	GAGCACAACCTTACCTCCCGAAGCTTCTTACAGACCTTCAAGAGCTTGAAGAA	2827
Qy	4211	ggaatcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	4267
Db	2828	GCAGTTTGAAGGCAATGACCCCTTGAAGCACTGTGAGGCGCTGAGCGGAGATGCC	2887
Qy	4268	aagcaacacagtcagcttccacagtgatcccaagcagcagcagcagcagcagcagc	4327
Db	2888	CATCAGCTACGTGCAACCTGGGCACTCAACTGTGATTTGTAGC---ACCACTTGG	2944
Qy	4328	gttthaagaatgaacaactcaataagcagcagcagcagcagcagcagcagcagcagc	4387
Db	2945	AGGAACGTAAAGTCTTAACTCACTCAACACCACTTCTCTCTGTGTGTGTGTGTG	3004
Qy	4388	atgtgcgcagggcctaagaagcagatgtgctgtgataactgtggggaactccacac	4447
Db	3005	GGTGTGTTCTGT	3063
Qy	4448	gagccat 4455	
Db	3064	GACCTTT 3071	

```
RESULT      5
AX149461
LOCUS       AX149461               3625 bp      DNA      Linear      PAT 08-JUN-2001
DEFINITION  Sequence 17 from Patent WO0136638.
ACCESSION  AX149461
VERSION    AX149461.1   GI:14347936
KEYWORDS
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3625)
AUTHORS    Shimkets,R.A., Lichenstein,H., Vernet,C. and Fernandes,E.
TITLE       Polypeptides and nucleic acids encoding same
JOURNAL    Patent: WO 0136638-A 17 25-MAY-2001;
            Curagen Corporation (US)
            Location/Qualifiers
FEATURES
            source
                1..3625
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT  896 a      971 c      914 g      844 t
ORIGIN
Query Match      34.0%  Score 1599.4;  DB 6;  Length 3625;
Best Local Similarity 74.1%  Pred. No. 0;
Matches 2189;  Conservative 0;  Mismatches 641;  Indels 125;  Gaps 8;

OY 1508 ccttcaatcagagcccgatataaagacttgatcagcagagccctcatgcccagcttc 1567
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 571 CCTCGATCTCTGTTAGAGTTAAAGACTGGACAAATACGGTTTATGGCCCCAGGTTTC 630

OY 1568 ttgcctatcacgtggtggtggtgcagcagctgctgtgtgagaaacctaaagtggaccacaa 1627
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 631 TTCGGTACCATGTGCTGCCCTCCAGCACAGCTGCTTCTGAAACCTGAAATTTGACTCAA 690

OY 1628 gtggcagcagccctcagaagagcagtttcacatctctgtctctcagacactgtgttca 1687
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 691 ATGCTACTTCCCTCCAAAGAGACCAATATGATCTCCGTCTCTCAGACAGCGGTATTA 750

OY 1688 taacaatgagcagaggtctctgtcagtgacatcatcagcacaatgctgctcatccag 1747
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 751 TAAATATTAAGGCTAAGATCATATCCAGTATATCATAGTACTATATGGATTTGTTCTATA 810

OY 1748 ttatagacaagctgctgtctcctccaaaactgtcttatcaccctccaaagatgcttgagca 1807
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 811 TCATAGACAAATTTGCTATCTCCCAAAATTTGCTTATCATCTCCCAAGACAACCTGTGAA 870

OY 1808 ggggtctgcaaaatcttactacagtgagcagaacccaagatatalaccaaatctgcagaagt 1867
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 871 GAATTTCTGCAAAATCTTAGACTTTGGCAACAACAATATGCTACATCAAAATTTTACCAACT 930

OY 1868 tgatcagagactcaggtctgtgtcagtcacactgactcatccacacccccagctcag 1927
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 931 TAAATACAGGACTTCAGGTTGCTGAGTGTGCATACCGATCCATCCACACCCAGTCATC 990

OY 1928 tctctgtgctcagcagacaagcccttgaaagccctgtgcccagagcagcagaaactctcgt 1987
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 991 TCTTCTGTGCCACCGACCAACGCTCCATGCCCTTACCTGCTGACCAACAGGACCTTCCTGT 1050

OY 1988 tcaatcagacagaacagcagctgaaagcttctacctgaaattccagagtgctcgaagct 2047
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1051 TCAACCAAGACAAAGACAGCAAGCTGAAGGAGTATTGAAATTTTCATGTGATACAGATG 1110

OY 2048 ccaagactttagcttcagcagctcccaaggtctgtcttctcgtgaaagcccttgaaagctcag 2107
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1111 CCAAGGTTTATGCTGTGATCTTCCACATCTCAGCTCCCTGGAAGACCTTGCAAGTTTACG 1170

OY 2108 agctgagtgtagagtgtagaagcagtgacatgcagtgagctcttcttaaacgaagcaaa 2167
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1171 AGCTGAGTGTGAATGTGGAGCTGGCAAGGACATCGGTGACCTCTTTCTGAATGGCCAAA 1230
```

```
OY 2168 tgtgagatltcataccgggagactctgtgttgagcgtgagtgtagcctatgagcattagct 2227
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1231 CCTGCAGAAATTTGTGACGGGGAGCTCTTTGTTTGAACCTGAGGTGTGGCTTACGACATTGACT 1290

OY 2228 gctactatcatgaatccctaccctaggtgagcagatgagacatttctactcctcagatc 2287
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1291 GCTGTGATTTGATGCCAACOCCTGGGGGCGGCTGTGACACCTTTACTACTTTTCATGGCTCT 1350

OY 2288 cggggaggtgcggaaagttgacatttcaactcccacaaatgcaccactgaaagcagaagc 2347
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1351 CGGGGAGCTGTGGAGCTGTGTCAATTAATCCAGCTGGCCCAAGGCTGAATCAACCAAGG 1410

OY 2348 ggcgtgaagaagaagtgtatctaacaccccgcttactcttgaagaagagcgtgaaagctc 2407
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1411 GTTGAACCAAGAGTGTCTTCAACAC--TGCCTTTCAAGAGGAACCTGGAAGGCTGCC 1467

OY 2408 agaaccctgtgacccgtgtgtatcccaaccccaagtgctgcatggttactactatgacag 2467
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1468 GGGAGCGGTGACAGCCTGGTGATACAGATCCCAAGTGTGCAAGGGCTACTTCGGGGCAG 1527

OY 2468 actgtcagagctgccttggaagaccagatataacgcgtgtatacaaacccgggcatgtgcg 2527
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1528 ACTGTCAAG---GGTGAAGGTCGCTCTTCCGCCCTCGCAACTCTAAAGCTGTCTGCTTGG 1583

OY 2528 atctgtacacaccctatggaagatgctcatatccacacccgcttcaacggagcagctcgt 2587
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1584 ATCACTACTTCGGCCACCCGAGAGTGTAAATATGCAACCGGCTTCAATGAGAGCGGTGTG 1643

OY 2588 agctctgtgcacatgagagatttgagcctgactgccaaccccgagctgctccagacatg 2647
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1644 AGATTCGTGGCGGGGAGATTGSGGCTGATTTGTCGCCCTGTGTGGCTGTACAGCCACAG 1703

OY 2648 gaacagtgtagaagggatltcaacaggtctccgggagtgccctgtgtgaaacaggttgagcag 2707
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1704 GACAGTGGAGATGATGGATACAGGGGCTCCGGGCACTGCTGTGAAACCGGGGTGACAG 1763

OY 2708 ccgcttggtgtgacatcccaagctgtatccgagtgtagcaacactgtgtcttcgctg 2767
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1764 GCCCTCTGTGTACCTACAGCAGATTTTGGCTTCGTGATGTATGAGCTCTCTTGTGTGTC 1823

OY 2768 agcccaactgtatgagagacacacacagctgtgtgtgtaacttgaacagaaagttagcagga 2827
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1824 ATGCCACCTGTATAGAGAGAACAAACAGCTGTGACTGTAACTGGATTATGAAAGTGTACGGA 1883

OY 2828 tcaatgtcacagctgtgactcttgcacaaacagaaacacgggggctgtgcgaagtgctcta 2887
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1884 TCACATGCACAGTGTGTGATTTCTGCAAAAGAGCAACGGGGGCTGTGCAAAAGTGTGCCA 1943

OY 2888 agtgtctccagaagaagcaccacagctctcttgacgtgtgcaagaaggtctacaagggagtg 2947
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1944 GATGCTCCCAAGAGGGCAGAAAGTCTCTGTGACGTGCCAGAGGATACAAAGGGGAGCG 2003

OY 2948 gctacagctgtcatagagaataagccctgtgtcagaagtgctcaacggggagtagccatgagc 3007
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2004 GGCACAGCTGTGACAGATAGACCCCTGTGTGACAGAGCGCTTAAACGAGAGGTGTACAGAGC 2063

OY 3008 acgccaactgtcagagatgacgggcccaggaagcatalaagtgtgaaatgaaagctactatg 3067
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2064 ACGGCACCTGTATAGATGTACAGAGCCCGGGCAAGCAACAAGTGTGAGTGTAAAGTCACTATG 2123

OY 3068 tctgggagcaggttgagactgtgagccttgagcagctgtgcgcttgacacgctgtgttaagagca 3127
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2124 TGGAGATGGGCTGAACTGTGAGCCGGAGAGCTGCCCATTTGACCGCTGTCTTACAGGACA 2183

OY 3128 acgagcagtgccacccagatgtccagctgtgtcagagacctactctccagacacgacagctag 3187
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2184 ATGGGCACTGTGCATGACAGCAGCAATATGTGTGACACTCCACATTCAGAGTATACACTGTGG 2243

OY 3188 gagtatltccatcacgcttcccactgggacagttacaacactgtacatttgaacaagccaag 3247
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2244 GGGTGTTCATCACTACAGCTCCACAGCTGGGCGAGTATTAACGTGACCTTGTGCAAAAGCCAGAG 2303
```



QY	3248	aagctgtgtgcaaaagaagctctgagccatctagccaccttaaacacgtctccctatgtccaga	3307
Db	2304	AGGCTGTGTGCAAGCAAGCTGGAGACCATGTGCAACTTACACAGCTCTCCATATGCCAGA	2363
QY	3308	agggcaagatctaacctgtgtctcgcccgctgtgctgagagatgvggcgggtttccataccga	3367
Db	2364	AGGCAAGTACCAACCTGTGCTGACGAGCGCTGGCTGGAGACGGGGGGGTGGCTTACCCCA	2423
QY	3428	ggggccaagaagatgtaaatgtgtgagatgtcttctgttaccgatatgaatatgtlaactgcga	3487
Db	2484	GACCCAAACAAGATGAAATGTGGAGATGTTCTTGCTATGCGATGTAAG-----	2531
QY	3488	ccctcaaggagctatgtgvgagatgtgtctctgtcaagtvggaacctgtcgaagctc	3547
Db	2532	-----	2551
QY	3548	tcatgtctctccctcgctctcaacaacttccctgacagaggtgtgtgtcttttccaagagt	3607
Db	2532	-----GAAGTGTGGCTATTCCAAAGCT	2556
QY	3608	cagcccggaagacagggcatlttttgaaacacctgtacctgtctccatcctgtgtgacacctgtc	3667
Db	2557	CAGCTTCAGAGCCGGTCACTTTTATGAACACCTTACCTACCTGTGCTATCGCGGGCACCTCT	2616
QY	3668	ttgtgcacagaagaagtvgggtctacccgggaataagaagctgtgtcgtgcggcgacattgac	3727
Db	2617	TTGTGCTCACAGAACAGTGGGCTGGGGGACAAATGAGACCTTGTCTGGGGGACATCGACG	2676
QY	3728	accacactactaatgtlcaacgtctcccttlttaaatgtacctgtgtcaatvgtaaccttctga	3787
Db	2677	ACCACCTCGCAATGTACAGATGTTTCTTACAAATGACCTTGTCAATGGACCAACCTGCG	2736
QY	3788	ggactatgtctvgggaagcaactgtctatctatccttaagccagagaccagttcaccc---aag	3844
Db	2737	AAAGAGAGCTGGGAAGCAACCTCTCATCACTGCTCCGACGAGGACCCATCTCCAAACGACGG	2796
QY	3845	agaccaggtctgttgaatgagaagatccatctctcgaatgagacatcatcgccgcacatgaa	3904
Db	2797	AGACCAGTTTGTGTAATGAGGAAGACCATCTGTGAGTGGGACATCTTTCCTCCAAATGGGA	2856
QY	3905	tccctcatatattctctgaaccccttgaagctctcccaagcgacaaagcgctgccact	3964
Db	2857	TCATTTCAATGTCATTTCCAGGCGCTTTAAAGACACCCCTGCCCTGTACCTTGACCCACA	2916
QY	3965	cttggccctggggaagagatattctctgcccgtgcctctgtgtcaactgtgtgattgtctcgg	4024
Db	2917	CTGCTCTGGGAGCGGGATCTTCTTTGGCATCATCTCTGTGACTGGGCTGTGTCTGTGG	2976
QY	4025	cagcttactcttactcttcgcgtctaaagaagcaaccaactgtttccagcgttttga-ctag	4083
Db	2977	CTGCTTACTCTTACTTTTCGGATTAACCGGAGAACATTCGGCTTCCAGATTTTGTAGTGG	3036
QY	4084	aagagacatgtatgtctgtgtgttttggcaagcagcagcccaagaatatcgcgaacctc	4143
Db	3037	AAGAGGACATTAATGTGTCAGAGCTCTTGGCAGACGAGCGCTGAGATATTCTGAAACCCCT	3096
QY	4144	tgatatgacctcaagcgccgagcaacccccagagatctccctgtgaacctctcaagaacctgt	4203
Db	3097	TGTATGAGAGCAACACTCGAGCTCCGCCACAMACCTTCTTCAGACACCCCTTACAGGACTGTG	3156
QY	4204	gagaaacagatctbgaggaagacagcaccctctgggggacatg---cggtctctcaataga	4260
Db	3157	AAGAGCGCACTTGAAGGCAATGACCCCTTGAGGACACTGTGAGGGCTTGACCGGGAGA	3216
QY	4261	agaccagaagaacacacagatcagagtttccacagttgattccacccacacagctgtcttcacag	4320
Db	3217	TGCCAGCATATCACTCTGGCACCCTGGGCCATTAACATGTGGAATTTCTCAAG---ACCAATT	3273
QY	4321	atcagttgtttaaagaatgacaacatctaaagccagccatacctcacaccttctgtgta	4380

Db	3274	GCCTTTTAGGAAGTAAAGTCTCTTAAGCACTCAGAAAGCCATACCTATCTCTGGCTG	3333
Qy	4381	atcttggaattctgcgccaggctaaaggacatctgtcctgatactcgtgggacctccac	4440
Db	3334	ATCTGGGGGGTGTTCCTGTGGTGAGAGATGTGTTG-CTTGCCACCACCACTAAGCTTC	3392
Qy	4441	ctccctgaacctat	4455
Db	3393	CTCCCTGACCCCTTT	3407
RESULT 6			
AX149459		2483 bp	DNA
LOCUS	AX149459		Linear
DEFINITION	Sequence 15 from Patent WO0136638.		PAT 08-JUN-2001
ACCESSION	AX149459		
VERSION	AX149459.1		GI:14347935
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2483)		
AUTHORS	Shimkets, R.A., Lichenstein, H., Vernet, C. and Fernandes, E.		
JOURNAL	Polypeptides and nucleic acids encoding same		
TITLE	Patent: WO 0136638-A 15 25-MAY-2001;		
FEATURES	Curagen Corporation (US)		
source	location/Qualifiers		
	1..2483		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	619 a 648 c 639 g 577 t		
ORIGIN			
Query Match			
Best Local Similarity	26.5%;	Score 1244.8;	DB 6; Length 2483;
Matches 1469; Conservative	80.5%;	Pred. No. 0;	
	0; Mismatches	352; Indels	3; Gaps
1:			
Qy	1508	cccttaatcatbagccccgagtaaagaactggaatcagaagggcctcatgtcccaagtc	1567
Db	571	CTCTGATCTTCTGTTCACAGGTTAAAGACTGGAGCAATAACGTTTAAATGGCCCCAGTTC	630
Qy	1568	ttcgcatacaagtgatgtggtctgcagcagcgtctgtttggaacaactaaagtaccacaa	1627
Db	631	TTCCGGTACCATTGTGTCCCTGCCACACAGCTCTTCTGGAAACCTGAAATTTGATCTCA	690
Qy	1628	gtgcacagaccctcccaagagaagccagttccatctctgtctcgaagacaatgctgtca	1687
Db	691	ATGCTACTTCCCTCCAAAGAGAGACCAATATGTATCTCCGTCTCTAGAGACAGGTGTATA	750
Qy	1688	taaaacaatbagcagaaggtccctctccagtgacatcatcagcaaccaatggtcatcaag	1747
Db	751	TAAACAATTAAGGCTTAAGATCATATCCAGTATCATCATGATCACTAATGGAGATTGTTCA	810
Qy	1748	ttatagacaagttgctgctcctccaaaacttgcttatcaacccccaaagaatgcttggtca	1807
Db	811	TCATAGACAAATGCTATCTCCCAAAATTTGCTTATCATCTCCCAAGCAACTCTGGAA	870
Qy	1808	gggtctcgaataattcttactacaagtgcaagcagaacaacagatataccaatlcagcaagt	1867
Db	871	GAAATCTCGCAAAATCTTACGACTTTGGCAACAAACAATGGCTACTATCAATTTATCAACT	930
Qy	1868	taatacagaactcaggtctgctgctcaagatcatcaatgactcaatccacacccagtcactg	1927
Db	931	TAAATACAGGACTCAGGTTTGCTGAGTGTATATCACCGATCCCATTCACACCCAGTCACTC	990
Qy	1928	tctcttgacctacggacaagaaccctctgaaagccttgccccagagaagcagacttctgt	1987
Db	991	TCTTTGGGCCACCGCCCAAGCCCTTCATAGCCCTTACCTCTGAACAACAGACTTCTGTT	1050
Qy	1988	tcaatcaagaacaagaagcaagctgaaagttcttacctgaagtccaagtgtatccgaagact	2047



Db	2128	TCGGAGATGGCGTGAACGTGTGAGCCGGAGACACGTGCCATTTGACCCGCTGTTCACGAGACA	2187
Oy	3128	acgagacatgcccacccacagatgagcagctgtgacagccttacttccagacagacagcgttag	3187
Db	2188	ATGGGCACTGGCCATGACAGACGCGCAATATGTGTGCAGCTCCACTTCACAGATATACCACTGTTG	2247
Oy	3188	gagatattccattacgctctcccccactgtggccagctaacaaatgacacatttgcacaagcccaag	3247
Db	2248	GGGTGTTCATCTACGTGCTCCACATGCGGCGCCAGTATTAAGCTGACCTTTGACAAAGCCAGAG	2307
Oy	3248	aagcctgtgcccagaagaagctgacacatagcaccactaaccaagctcttctatgcccaga	3307
Db	2308	AGGCTGTGCCAACCAACCTGGACCAATGGCAACCTACACCAAGCTCTCATGCCAGAGA	2367
Oy	3308	aggccaagatcacctgtgctcg 3331	
Db	2368	AGACTGGTATTCTTTTACCAAG 2391	
RESULT	7		
LOCUS	AX149455	2512 bp	DNA linear PAT 08-JUN-2001
DEFINITION	Sequence 11 from Patent WO0136638.		
ACCESSION	AX149455		
VERSION	AX149455.1	GI:14347933	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2512)		
AUTHORS	Shimkets, R.A., Lichenstein, H., Vernet, C. and Fernandes, E.		
TITLE	Polypeptides and nucleic acids encoding same		
JOURNAL	Patent: WO 0136638-A 11 25-MAY-2001;		
FEATURES	Curagen Corporation (US)		
source	Location/Qualifiers		
	1..2512		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	623 a 657 c 651 g 581 t		
ORIGIN			
Query Match	26.4%; Score 1243.2; DB 6; Length 2512;		
Best Local Similarity	80.5%; Pred. No. 0;		
Matches 1468; Conservative	0; Mismatches 353; Indels 3; Gaps 1;		
Oy	1508	ccctcaatcatgagcccgagatlaaagactggatcagcagagcctcatgtcccaagttc	1567
Db	600	CCTGTGATCTTCTGTTACAGGTTAAAGACTGGAGCAATATACGGTTAATGCCCCAGGTTTC	659
Oy	1568	ttcgcattacagctgtgtgctgcgcagcagcagctgcgtcttgagacaacctaaagtgcaccaa	1627
Db	660	TTTCGGTACCAGTGTGGTGGCTGCCACGACGCTGTTGGAAGAACTGAAATTTGATCTCTAA	719
Oy	1628	gtgcacagacccctcaagagagagcaggttccatctctgtctctcagaagcagctgtgtca	1687
Db	720	ATGCTACTTCCCTCCAGAGAGACCAATAGTCATCTCCGTTCTCTAGAGACAGGTTGTATA	779
Oy	1688	taaaatagtagcgaagttcctgtctcagtgacatcagatcagacaacatgagctcatccagc	1747
Db	780	TAAACATTAAGGCTTAAGATCATATATCCAGTGTATATCATAGTAAGTATGGATATTGTTATTA	839
Oy	1748	ttatagacaagtgtcgtctcccaaaaactgtctatcatcaccaccaagaatgccttgagca	1807
Db	840	TCATTAAGACAAATTTGCTATCTCCCAAAATTTTGCTTATCTACCTCCAAACAAACCTCTGAA	899
Oy	1808	gggtctgcacaactcttactacagctgtagcagaacaacacagatatcaccaattcagcaagt	1867
Db	900	CAATTCTCTCAAAATCTTACGACTTTGGACACAAACATGGCTTACATCAAAATTTTACCAACT	959
Oy	1868	tgatacagagctcaggtctgtctgtcagatcattcactgattcatcatccacccccagctacgt	1927

Db 960 TAATACAGAGCTAGGTTTGTCTGATCTATCCACGATCCCATCCACACCCCACTCATC 1019  
Qy 1928 tctctgacctagagacaagcccttgaaagccttgcccccaagacaagacttctgt 1987  
Db 1020 TCTTTCTGGCCACCGACCAAGCCCTCATGCTGCTGCAACACAGGACTTCTGT 1079  
Qy 1988 tcaataagaacaagaagaagctgaagcttctacctaagcttccagtgatccgagact 2047  
Db 1080 TCAACCAAGCAACAGACAGACAGTGAAGGATTTGAAGTTTCATGTGATACGAGATG 1139  
Qy 2048 ccaagacttagcttaagactccccaagctctgtctccctgaaagacctgaagctcag 2107  
Db 1140 CCAAGGTTTATGCTGTGATCTTCCACATCCACTGCTGAAACCCCTGCAAGGTTGAG 1199  
Qy 2108 agctgaagtgtgaagggtgtggaagctgacatcaggtgagctcttcttaagaacaa 2167  
Db 1200 AGCTGAATGTGCAATGTGGAAGCTGGACGGACATCGGTGACCTTTTCTGAAATGGCCAA 1239  
Qy 2168 tgtgcaattcaatacaacccggagactctgtttgaagtggtgtgtgactatgact 2227  
Db 1260 CTTGCAAAATTTGTGACGGGAGCTTTGTTGACCTGGGTGCTGACGCAATGACT 1319  
Qy 2228 gacctaccatgaatcccaacttagtgccgaatgtgacacttctactactcctgatatc 2287  
Db 1320 GTCTGCTGATTGATCCACCTGGGGGCGCTGTGACACCTTTACTACTTTTCATGCTCT 1379  
Qy 2288 cgggggagtcggaagctgtgacttcccaatcccaatcccaactgaagacaagcaag 2347  
Db 1380 CGGGGAGTGTGGAGAGCTGTCTCAATCTCCACTGCCCCAGGTGGAGTAACCAAAAG 1439  
Qy 2348 gcgltgaagaagaagtgtatatacaacccgttactcttcgaagagaagctgtgaagctg 2407  
Db 1440 GTGTGAAGCAGAAAGTGTCTCTACAAAC---TGCCCTTCAAGAGCAACCTGGAAAGCTGCC 1496  
Qy 2408 agaaccgtgtgacccgtgtgtatccaaaccccaagctgtgacatggttactatctgca 2467  
Db 1497 GGGGCGGTGTGAGCTGTGTATGATGATATCCCAAGTGTGTGCAAGGGCTACTTCCGGCGAG 1556  
Qy 2468 acttcaagccctgtgcttggaagacagatacaacccgtttaaacaacgggagctgtgccc 2527  
Db 1557 ACTGTAGGCGCTGCTGTGGAGAGCAAGTGGCCCGTGTAAATACCGGGGTCTCTGCCCTG 1616  
Qy 2528 atctgtacaacccaatggaagctgtgctatgtccacacccggttcaacgggagcaagctg 2587  
Db 1617 ATCAGTACTGGCGACCGGAGAGTGTAAATGCAACACCGGCTTCAATGGGAGCGCGTGTG 1676  
Qy 2588 agctctgtgcaatggaagattgtgacctgtatcagccccgaagctgtgtccgaagctg 2647  
Db 1677 AGATGTCTGTGCGCGGAGATTTGGCTGTGATGTGTGCTGTGCTGCAAGCACAG 1736  
Qy 2648 gacagctgtgaagggagatacaagctcccgagggtgcctctgtgaaacaaggtggaag 2707  
Db 1737 GACAGTGTGATGATGATCGATCCGGGCTCCGGGAGTGTCTTGTGAAACGGGGGTGGAGAG 1756  
Qy 2708 ccgctctgtgtgacactcccaagctgtatctgcaagtggtgacacactgtgtccgtg 2767  
Db 1797 GCCCTGTGTGACACTCAGGCAAGTATTGTCTGCAAGTGTGACCGCTTGTCTGTGCTC 1856  
Qy 2768 acgcaactgtgaaggaagaacaacggtgtgtgtgtaacttgaactaogaaggtgaagga 2827  
Db 1857 ATGCGACCTGTGAAGGAGAACACAGTGTGAGTGAATGATGATTATCAAGGTACGGAA 1916  
Qy 2828 tcaatgtacagctgtgactctgtcaaaacagaacaagggggtgtgtgcgaagctgtgta 2887  
Db 1917 TCAATGTGACAGTGTGTGATTTCTGCAAAACAGACAAAGGGGCTGTCCAAAGGTGGCA 1976  
Qy 2888 agtgtcccaagaagaagcccaagctctgtgtgcaagctgtgaagagcttacaaggggtga 2947  
Db 1977 GATGCTCCCAAGAGGACGAAAGGTCTCTGACGTGTGCAAGAGGATATCAAAAGGGGAGC 2036  
Qy 2948 gctcaagctgtatagatagataccctgtgtgacagcgtgtgtcaacggggaggtgcaatgagc 3007  
Db 2037 GCGCAGAGCTGCACAGAGATAGACCCCTGTGTGACAGAGGCTTAAAGGAGGCTGTACAGAGC 2096

Qy 3008 acgccaactgtcaagatagcggcccaagcaagcaataagtgtgaaagtacatcatg 3067  
Db 2097 ACGCCACTGTAAATATGACAGAGCCCGGCAAGCAAGGTGTGATGTAAAGTCAATATG 2156  
Qy 3068 tcggggaaggaagtgtgactgtgaagctgtgaagctgtgcgctcgacgctgttctacagga 3127  
Db 2157 TCGGAGATGGGCTGTACCTGTGAGCCGAGCAGCTGCCCATTTGACGGCTGTTCAGGACA 2216  
Qy 3128 acggaagctgtccaccagaagtcgagcgtgtgacacccctactctcaagacagacagctag 3187  
Db 2217 ATGGGCAAGTGTGATGACAGAGCCCAATGTGTGACCTTCACCTTCAGAGATACACTGTG 2276  
Qy 3188 gagtatccatctacgtccccaactgtgccaagtgaactgaactgtgacaagcaag 3247  
Db 2277 GGGTGTTCATCTACGCTGCCCATCTGGGCGAGTAAAGCTTGTGACAAACCGAGAG 2336  
Qy 3248 aagcctgtgccaagaagctgtgacacatagcaacacacacagctctctatgtccaga 3307  
Db 2337 AGGCTGTGCGCAAGAGCTGCGACATGCGCAACCTACAAACAGCTCTCTATGCCAGA 2396  
Qy 3308 agccaagatcacctgtgtcgtg 3331  
Db 2397 AGACCTGTATTCCTTTACCAAG 2420

## RESULT 8

AX149447  
LOCUS AX149447 2011 bp DNA linear PAT 08-JUN-2001  
DEFINITION Sequence 3 from Patent WO0136638.  
ACCESSION AX149447  
VERSION AX149447.1 GI:14347929  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2011)  
AUTHORS Shinkets,R.A., Lichenstein,H., Verneuf,C. and Fernandes,E.  
TITLE Polypeptides and nucleic acids encoding same  
JOURNAL Patent: WO 0136638-A 3 25-MAY-2001;

Curagen Corporation (US)  
FEATURES  
Location/Qualifiers  
source 1..2011  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 458 a 551 c 583 g 419 t  
ORIGIN

Query Match 25.4%; Score 1196; DB 6; Length 2011;  
Best Local Similarity 77.6%; Pred. No. 0;  
Matches 1524; Conservative 0; Mismatches 330; Indels 110; Gaps 2;

Qy 1868 tgalacaggaactcaggtctgtctgacatcactgaactccatcaaccccaagctcag 1927  
Db 158 TGTTCAGAGACTCAGGTTTCTGTGATGTATCACCAGATCCATCCACACCCCACTCATC 217  
Qy 1928 tctctgtgctcaaggaacaagccctgtgaagctgtgcccccaaggaagcaagacttctgt 1987  
Db 218 TCTTCTGTGCGCCAGCCCAAGCCCTTCATGCTTACCTGTGTAACAAAGGACTTCTGT 277  
Qy 1988 tcaataagaacaagaagaagctgtacatcttactgaagcttcaacgtgtgacagact 2047  
Db 278 TCAACCAAGAGAACAGACAGCAAGCTGAAGAGTATTTGAATTTTCATGTGATACGAGATG 337  
Qy 2048 ccaagcttagctcagacctcccaagctgtgtctctgtgaagaaacctgtcaagctcag 2107  
Db 338 CCAAGGTTTATGTGTGATCTTCCACATCCACATGCTGGAAGACCTGTGCAAGGTTTACG 397  
Qy 2108 agctgaagtgtgaagtgtgtgaagctgtgacatcagctgtgtgactcttcttaagaagaa 2167  
Db 398 AGCTGAGTGTGAATGTGAGACTGTGACAGGACATCGGTGACCTCTTCTGTAATGGCCAAA 457

OY	2168	tgcagatcacaacaccgggaacccctgcttgcagctgggtgtgtgacctatgcatgcatc	2227
Db	458	ccctcacaatttggcagccgggagacctgtgttttgacctgggtgtggccctacggcattggact	517
OY	2228	gcccaccatgaatccctaccctbgtggtgcgtatgtgaaccttactactccctgatatcc	2287
Db	518	gtctgcctgatttgatccacaccttgggggccgctgtgcaccttgcctactcttgcattgctc	577
OY	2288	ccggggagatgcggaagcttgcatcttccaccccaatggccacttgaaggacagcccaag	2347
Db	578	cgggggagctgtggagacctgtgtcaatattctccacgctggcccaagtgtagaaccannaag	637
OY	2348	gcgtgaagaaagatgtatctacaacccgttacccttcagaaggagacgttggaagctgac	2407
Db	638	gttggaagcagaaatgtctcttaccamcc---tgccctttcaaggagaacctggaaagctgcc	694
OY	2408	agaaacctgtgacccgtgtgtatcccaaaacccccagttgctgcctatggttacttcatgcca	2467
Db	695	ggggagccatgcagcctgtgtatcacaatcccccaggtgctgacaagggctacattcgggcgag	754
OY	2468	actgtacagccctgcctccggggagacagatacaacccgtgtataaacacggggacatgtccgc	2527
Db	755	actgtcagagcctgcccttggagagcaccagatgccccgtgtatattaccgggctgtctgccttg	814
OY	2528	atctgtacacaccatlyggaacgtgtgctctatgtccacacccggtctcaacgagacagctgcg	2587
Db	815	atcagctactgcggccacggagagctgtatattgcamaacccggctttcaantggagccgctgtg	874
OY	2588	agctcctgtcgtacgtggaagatlttgggctgtactgtcagccccgcagctgtctccgagcatg	2647
Db	875	agatgtctctggccggggagatttgggcccgtgattgtctgcctctgtgcctgtacagccacgc	934
OY	2648	gaacagttgatagaaggagatcacacagctccggggagatgtgctctgtgtaaacagatgtgaag	2707
Db	935	gacagtgctgatgatgatgatcattacgggctccgggcagtgctctgtgtamaacggggtggacag	994
OY	2708	ccgcttcgttgaactccccaacagctgtatctgcaggtgtgacaacctgtctgtccctgtgc	2767
Db	995	gcccctcgtgtgaacactcagggagctgtttgtgtcgaatgtgtacgctccttgttctgtctc	1054
OY	2768	agccgccctgtacggagaaacaaacgtgtgtgtgttaacttgaactagaaagtgtacgga	2827
Db	1055	atggccacctgtaaagaaamaacacacgtgtgagctgtatcctggatttatagctgacgama	1114
OY	2828	tcacatgcacagctgtgtgaacttctgcaaaacaaacaaaggggcgtgtgcgaagtgcgta	2887
Db	1115	tcacatgcacagctgtgtgtgattttgtgaaacagacacacggggcgtgtccaaagggtggcca	1174
OY	2888	agtgtcccaagaaagagcaccagaactctctgcagctgcgaagaagagctcacaaaggagatg	2947
Db	1175	gatctctcccaagaaaggacacaaagctctcctgcacgtgcccaagaaaggatattcaaaaggagac	1234
OY	2948	gctacagctgtcatagatagatagaccctgtgcagacggtgtcaacggggatgtgcatagac	3007
Db	1235	ggcacacactgcacagaaatagacctctgtgcadacgcgctttaaacgagagctgtcacagac	1294
OY	3008	agccgacctgcagatgtacggggcccgagcaagcatatagtgttaatgtataaagtcatcatg	3067
Db	1295	acggccacctgttaaatatcacaggcccgcggaagcacaagtgtagtgaagtcatctatg	1354
OY	3068	tcggagagcgaagtgtgacttgcagctgcgaagcgtgcggctgcagcgtgttacaagga	3127
Db	1355	tcggagatgggctgtaaactgtgtgaccggagacactgtccatttaccgcgtgtttacagga	1414
OY	3128	acggacagtgccaccagatgtccagctgtgtgcagaccttacttccagagacgcagctag	3187
Db	1415	atgggcagctgcacagcagagcccaaatgtgcgacacttcacattccagagattaccactgttg	1474
OY	3188	gaagtattccatctacgtctcccaactgtggcccgatataacatgtacatlttgacaagccaaag	3247
Db	1475	gggtgtttccacttaccgtctccacctgtggccagattatgaactgtgaccttttgcacaaagccaaag	1534

QY	3248	aagcctgtgtc	aaagaagctctgc	accataagc	actctc	ctactgtcc	caga	3307
Db	1535	AGGCTGTGTCC	CAACGAAGCTGG	ACCATGG	CAACCTAC	CAACAGCTTCT	TGTGCCA	1594
QY	3308	aagcccaag	tataactctgtc	ctgcgcgcgc	ctgtc	ctggaag	tggcgggctgtgc	ctaccga
Db	1595	AGGCCAAGT	ACCACTGTGCT	TACGACAGCT	TGGCT	TGAGACCGGGGG	GTGGCTT	ACCCCA
QY	3368	ctacgtatg	cctcctctaga	agtg	tggtcaaa	cgttctgt	ttgga	tcg
Db	1655	CAGCCTTG	CGCTTCC	CAGAAC	CTGTGGCT	GTGGTGTGG	TGATGTG	AGCTATG
QY	3428	ggcgcaaca	agaagtg	aaatgtgtg	agctgtc	ctctgt	cttacc	ggatgaaga
Db	1715	GACCCAA	CAAGAGTGA	AAATGTGG	ATGTCTT	TGCTATG	AGATGA	AAAG
QY	3488	cctgaagc	agcagctatg	ggagagatg	tgctctc	gtg	taag	tgggaac
Db	1763	-----	-----	-----	-----	-----	-----	-----
QY	3548	lcatgtc	ctccctc	gtc	ctacaa	acttc	ctgacag	agtgctgtc
Db	1763	-----	-----	-----	-----	-----	-----	-----
QY	3608	cgagccg	agagaagc	atltt	tgaaac	acac	ctgtac	gtgac
Db	1788	CAGCTCG	AGCGCTGC	ATTTCTT	TGAAAC	ACTTGAC	TGACCTGTGC	ATCCCGG
QY	3668	ctgtgc	caacaaga	acagtg	ggcgtac	cgcggga	aaataag	ac
Db	1848	TTGTGCT	CCACAG	AACAGTGG	CGCTGGG	GGAGAA	TGACAC	CTTGCTGTG
QY	3728	accac	ctcaacta	atgtc	aacg	tctc	cttt	tta
Db	1908	ACCACCT	CGCCCA	NTCTG	ACGATG	TTTCTT	CTACAA	TGACCTTGTCA
QY	3788	ggacat	ctgtgga	agccaac	gtcgc	atc	actt	ac
Db	1968	AAACGA	GCTGGG	AGCAAG	CTGTCT	ATAC	TGCGCA	CGCAGG
RESULT 9								
AX149449		AX149449		1804 bp	DNA	linear	PAT 08-JUN-2001	
LOCUS		Sequence 5 from Patent WO0136638.						
DEFINITION		AX149449						
ACCESSION		AX149449.1		GI:14347930				
VERSION								
KEYWORDS								
SOURCE		human.						
ORGANISM		Homo sapiens						
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
TITLE		1 (bases 1 to 1804)						
JOURNAL		Shmukers, R.A., Lichenstein, H., Verner, C., and Fernandes, E.						
FEATURES		Polypeptides and nucleic acid encoding same						
source		Patent: WO 0136638-A 5 25-MAY-2001;						
		Curagen Corporation (US)						
		Location/Qualifiers						
		1..1804						
		/organism="Homo sapiens"						
		/db_xref="taxon:9606"						
BASE COUNT		426 a	487 c	521 g	370 t			
ORIGIN								
Query Match		21.6%;	Score 1016.4;	DB 6;	Length 1804;			
Best Local Similarity		82.0%;	Pred. No. 2e-279;					
Matches 1183;		Conservative	0;	Mismatches 256;	Indels 3;	Gaps 1;		
QY	1868	tgatacag	gactcag	gctgtc	gtctgtc	atgcatc	atgactc	atccacacccag
Db	158	TTTTCAC	AGACAC	AGTTTGT	GTGAC	TGTGAC	TGACAT	CCGATTC



GOHLEEDRTQOTRRMWTLAGOEITVTFNOFTKYSYKXKDOPOQTFNJKANNIANG  
VFHVYTGRLMOPASGTPGDPKRTIGOLIASTAEAFRFTLLENCGLPILDDGPFTV  
FAPSNEAVDSLRSDGRLILYFAGLSKLOELVRIYNNGLTVELIKSGRLITMANO  
VLAVNISPEGRILLGREGVPIORVDMYMANIYHMLDILPLPTLILPKRSEOH  
KIVAGSCYDCALNTSTCPNSVKIDITPKECVITHDPTGLVNLKGCASCTN0TIME  
QGCCGFEGPDCTOCPGFSNPKYKNCSDGIDQNGACILCPDKGLACHICSNPMK  
HGEOQEDCGGVHGLCDNRPSPGSGVQDGTGAPGSPFCNMSMDCCPTGLAOHCL  
HARCVOBVARCRLDGFEDGFSCTSPNPSHDPGSCNACVCSGLDTHHCTC  
HKMSGDRGVCAIDECLELDVRCGCTHATCSYRPOSRCTKCFGFGDQVCSPID  
PCBANGSGCHLELEANAHSFYOMLKSACTIYPADRPTALVSEAVRSLSEDR  
AFWLOPRLPMLVRAHFLDQALFEELRLUGQEVATLNPTRWELRINTSGRWONA  
SVDVADLILATNGVLHIIISOVLLPRGDPVGGGILLQOLDVPAFSLRRELLOHGLVP  
OIEATATYITEVPTNRSLEAOGNSHLLADVVRHHVILGEALSMETLKGHRNLG  
PAHMTVFYHNSGOPEVNHVPLEGPMLEAPRGLICLSVLTVGSSNCSHAEALREK  
CVNCTRRRCRTOGPOLODTPRKSVCYRSCSTYCAKKTOPVCCPEPRTIC  
EPCPGJGVCVSGHOCODRPLGSECHCEFHCTAEVCELGKRYNCTVCCAH  
GLCOBLOGDSCVSNVGMQIRCDQKTTSPCKPCRCODNMVDSAGASTCACAG  
YSGNCFECSVDPCAHGSGSPHANCYKVAAPGQRTCCODGMDGELCOEINSLI  
HHGGCHIAECITPQPOVSCSREGVSGDITRICELDPSCKNNGGSPYATCKSTG  
DGORCTCDTAHTVGDGLTCRAVGLLELRKHASFSLRLEYELKGDGFTTFVP  
HADMSNLODELARIRAHROLVFRYHVYGRSLSEDLLEGVATLSDGHLRFRSE  
EGSTYLDNFARVYSSDHEAVNGILHFDIVLLPRLALHMPEDDAITPRNVTAAOGF  
CYKIFSGILKVAGLPILREASHRFTLMPDIAFRALPDRQWMLRHEHRLDLAA  
ILRGHMTNRNVALASDLPNLGLPRLTMBGTPISFCSRTPRGELMWGEDDARIVORHL  
FEGGLAYGIDOLEPPLGARGARDHETRPLRLNTCSICGLEPPECBEGSOEOPBACW  
RFYKFTWSPPLHSLGLRSVWVHPSLMRGRTGLGRHRCITWTMKFSCGHHGSE  
COACGGSPSPCDRGCYMDGMSGOCILCSGFTACELCAPAGFHHOCACCTV  
HGRDEGLGSGSCDCDEGMPRCENOLEIOPLEIOPLEIOPLEIOPLEIOPLEIOP  
GDRXYCTAADLCODGSGCSEHANCISOYGTWYTCCLPDYEDGSCSRKNCCTTGH  
GCSSEHANCSTGLNTRRCECHAGVAGQLOCLESEPPVBDICLOPPCHSDAMCTD  
LHPEKRAVGHPLVATSGVPLNESEAEAEADQAVLASFOLSAADOLHNLICMG  
WLANGSTAHVPVPAVDCNGRVGLVSGARCNLESEMDAYCFRVDYADRCRNFVG  
DGIETNCKLKLVDLAATANFSTFYGMLGYANAPRGLDFLDFLDFLDFLDFLDFV  
EGFDNMTLGSDPLLEHASNATLISANSOCKLPAHSGLSLITSDADPNSWAPVA  
PGTYVYSRIIYWDIMAFNGIITHALASPLIARPOPVLAPEAPVAAGAVATAALL  
GLVAGALYLRARGKPMGFSGFSAFOAEDDADDFSPWEGTNTLVSVPNPFSGDTFC  
EPFDDSLLEDFPDTGRITLYK"

BASE COUNT 1234 a 2129 c 2099 g 1315 t  
ORIGIN

Query Match 14.8%; Score 698.8; DB 9; Length 6777;  
Best Local Similarity 51.2%; Pred. No. 2.4e-188;  
Matches 2094; Conservative 0; Mismatches 1867; Indels 125; Gaps 15;

QY 25 ctggaggaatgcccgacattccatttccgaggttacaattatattacaaccctggca 84  
DB 2333 CTGGACCTTGCTGCTGCTTCAAGCCCTTTCGGGAATTCCTCCAGACCATGGTTGGTG 2392  
QY 85 agtgaatcgaatctgcagatgcttatactgtgtctgtgcaacaacaatgaagccatcgaa 144  
DB 2393 CCCGAGATTGAGAGCTGCGCACTGCTTACACCATCTTTGTGCCACCAACCGCTCCCGGAG 2452  
QY 145 aactatataaggagaagaagaacacatctctaagaagatatattacggtatacgtatg 204  
DB 2453 GCC-----CAAGGCAACAGCAGTCACTTGAGGACAGACAACTGCGGACCACTGTGG 2503  
QY 205 gtccctgggggaaaagctcctctgaagaatgactgtgcaataagcgatgacaggaagaccatg 264  
DB 2504 GTCTGCTGGGGAGGCGCTTCTCCATGGAACCCCTGCGGAAGGGTGGAGACCCGCAACTCCCTC 2563  
QY 265 ctggaggttcttaactctcttcttcttcttcgcgaatgacgaagctgtatgtaatgaa 324  
DB 2564 CTGGGCGCCCGCCACGATGCTCTTCTTCAACACACAGTGGCCAGCTCAAGGTGAACCAT 2623  
QY 325 gctccaataaactacaacaatgtgaccatgataaaggagtgatccatggtcttgagaana 384  
DB 2624 GTCCTGACTGAAGGCCCTTGTCTGTGAGGCGCCCTGCGCTGATGTGTCGTGGGG 2683  
QY 385 gtctctggaataatcagaagaacagatgtgacaataatgacaacctatctgtgagaaggag 444  
DB 2684 GTCTGTACGCGTCAAGTGTGCTGCTGATAGCCACGCTGAGGCGCTTGGCGGAGGAAA 2743

QY 445 tctggaagaatgttccccaagaagccccctgcccacactgagacaaacacacttagagagc 504  
DB 2744 TGTGTAAACTGCGCACGAGGAGATTCCGCTGCACTCAGGGCTTCCAGCTGCAGGACACACC 2803  
QY 505 aggaataatcatctatccatctacttcaatcatalgaggaagagatccgatattcatcgggtgccg 564  
DB 2804 AGGAAGAGCTGTCTTCAACGATGTGCTT-----TTCCTTCTCCCGGGGCTGCTCT 2854  
QY 565 ccaacatgtgtggaacccaatcattacaagaagcctgtgtg-gctgtctcttcttgccaca 623  
DB 2855 TACACATGTGCAAGAAAGATTCACAGTGCAGGACTGTGCTGCTGCTTCTTGTGGCACGCTG 2914  
QY 624 tgcgaagcctgcccgcgggagaggttcaaaatgtgtgtctctgggaagcgctctgtctggga 683  
DB 2915 TGTGACCACTGCCACAGGGGCTCTAGGGGGGCTGTGCTCAGGCCATGGCAGTg-CCAGGA 2973  
QY 684 cgtgtgaaatgagcatcgtgcaagctgtgcaagctgtgagccttcaatgagacagcctgtga 743  
DB 2974 CAGCTTCTGCGGACAGGGGAGTGCACATGCCACGAGGCTTCCATGAAACGCGCTGTGA 3033  
QY 744 aacctgacatgaggagggaagtatgtatgatactgtgcaacaaagcatgtctctgtgtccatg 803  
DB 3034 GGTGTGTAGCTGTGGCGCTACGGGCCCAACTGCAACCGAGTGTGTGACTGTGCTCCATGG 3093  
QY 804 gagatgtgaacaaaggaaaccttgggaagcgctctctgtgaactgtgagctcgctcgagag 863  
DB 3094 GCTGTGCCAGAGGGGCTGCCAAGGGAGCAAGCTGTGTCTGTAACTGTGGGCTGGGAGGG 3153  
QY 864 agtgaagctgtgacatgagatcaacaacagacaactgtgcaacggagcctgtcaacacagtg 923  
DB 3154 CCGCCCTGTGACCAAGAAATATCCAGCCGCTTCACTGCTCCTAGGAATGCGACCCCAATGC 3213  
QY 924 caactgctcttggatcccaagcggaagaaacctgtgcaaatgtgagcgagatctcgagag 983  
DB 3214 CAACCTGTGACGAGTCTGCGCGGACCTTCACCTTCCGCTGTGCTGCGGGATACCTCCG 3273  
QY 984 gaatgaaacggtctgcacagcaatcaatgctctgtgagacgaagatgagatgttctac 1043  
DB 3274 CAATGCAATCTTGTGTTCAGAGSTGAGACCTTCCGCGCAAGGCGCATGGGGCTGCTCCC 3333  
QY 1044 aaaagccgaactgtlaaagaacaccccaaggaaacgggtgtgtgtgtgtaagagcaggtcta 1103  
DB 3334 TCATGCGCAACTGTACCAAGGTGCGACCTTGCGGCGAGCATGCACTCGCAGGATGTGTA 3393  
QY 1104 taccggcgagcgatcgtgtgtctgtgaaatcaaccggtgttgggaacaaatgtgtgcg 1163  
DB 3394 CATGGGCGAGGGGACCTGTGCGAGGAATTAACAGCTGTCTATCCACCAACGCGGGCTG 3453  
QY 1164 tgacagaatgtcagatgtcacacagacagggcccaacgaagcgctgtgtaactgtctgc 1223  
DB 3454 CCACATTTACAGCGCGATGCGATCCGCCACTGCGGCCACAGAGGTCTCCTGCAAGCTGCGGTGA 3513  
QY 1224 gaagtacactgtgagatgaaa---aggtctgtgtgttataatcgaatgtctgtcctaagaca 1280  
DB 3514 GGGTTTACAGCGGGGATGCGCATCGGACCTTGGGAGCTCTGGGAGCCCTGCTTAAAGAA 3573  
QY 1281 tggcggtcgtcagttccatttgccttctgtcaactacactgtggaagaatgaaagatatgtac 1340  
DB 3574 TGGAGGATGCACGCCCATATGTCACCTGTGCATAAAGACACAGGGGATGGGACACAGCATGTG 3633  
QY 1341 ctgcaagccgaactacacac---gggtatgtgaatcgtctgcgggggacagatctacagga 1397  
DB 3634 CTGCGACACAGCCCAACCGCTGGGGGACCGCTTCCACTGCGGTGCGGCTGCGGCTGGA 3693  
QY 1398 gcttcccaagaaccccttgaagcttccagtaacttctccagttgcaaggaagatgtgtccg 1457  
DB 3694 GCTCTGAGAGGATTAAGCATGTGCTTTCACGCTTCCGCTCTCGGAATATA-----A 3747  
QY 1458 agagcttctgtgacccctgagcccttcaacggtgtgtgcgacctgtgtctagcttccctta 1517  
DB 3748 GGAAGCTCAAGGGGAGATGGGCTTTCACCATCTTGTGAGCGGACAGCATTAATGAGCAA 3807  
QY 1518 tgagcccgatataaagactgtgatacagagcctccatagtcttccaggttcttctgataca 1577

Db	3808	CTGTCCAGAGATGAGCTGAGCCCGGATTGTGTGGCATGTGCCACGCTGGTGTTCCTACCA	1617
QY	1578	ctgtgtgagctgcgcgaagaagtgtctgtttgacaacctaaagtgcacaaatgctgcagac	1637
Db	3868	CGTGTGGTGTGGCGGCGCTGGCGAGCGAGAGACCTGTGTGAGCGAGGGGTACGCCACGCG	3927
QY	1638	ccctcaaggaggaagcagtttccatctctctctctcaagaacctgtgttcataaacaatga	1697
Db	3928	CTCTTCAGGGGACACCACTGGCGTTCAGCGAGAGGGAGGGGACGCAATATACCTCAATGACTT	3987
QY	1698	ggcgaaagtctcttccagttgacatactcagacaacatggcgttcaltccacgttatagacaa	1757
Db	3988	CGCGCGGTGTGTGAGCAGCAGACCATATGAGGCCGTGAACGGGCAATCTGCATCTTATTAACG	4047
QY	1758	gttgctgtctcccaaaaacttgcttatcaaccccccaaatgactgttggtgagtggttcgca	1817
Db	4048	TGTCTCTCTGCTCCCGCCGAGGCGCTGTGACATGAGGAGGAGGCTGTATGTCTCCATCCGAGAGG	4107
QY	1818	aaacttactacagttgacgaacaacagatataccaacttcaagtaacagtgtgatacaga	1877
Db	4108	AAATGTCACGCGCGCGCCGCCAGGCGCTTGCGTTTACAAAGATCTTCACAGGGCGCTCGGAAGT	4167
QY	1878	ctcagctgtgctcagtaactaactgaactccatccaaccccagctcaatgtctctctgccc	1937
Db	4168	GGCGCGGCTCTGCGCCCTGTGTTGAAAGGCAATCCATAGGCGCTTCACAAATGCTGTGGCC	4227
QY	1938	tacgagacaagacctggaagacctgtgccccagagacagagacgtcttcgttcaataaga	1997
Db	4228	CACAGACGCGCGCTTTTGAGACTCTGCGCTCCGGATGCGCAGGCGCTTGCTTACATGAGA	4287
QY	1998	caaaaagacaagctggaagcttactctactgaagtccaatgatacgtgatactccaagcttt	2057
Db	4288	CCACCGTGAACAACCTAGCAGCACCAATCTGCGGGGGCCACATGATTCGCAATGTGAGAGCTT	4347
QY	2058	agcttgaagactccccaagtgctctctcccggaagacctggaagcttaagagctgaagtggt	2117
Db	4348	GGCATCTGACCTGTGGCCAACTGTGGGCCCACTGTGAACCAATGATGAGGAGCCCACTCTCTT	4407
QY	2118	gaagtgtggaacttgcagtacatcggtgtagctcttcttcaagacaacaatgtgcagatt	2177
Db	4408	CTCTCGCAGGC---GAACGCGGGCCGGTGAAGCTCATGGTGGGTGAGATGATGCTCGCAT	4464
QY	2178	catacacccgggaactctgttttgaagtggtgtgtgacctatgtgcatgtgactcactacat	2237
Db	4465	TGTCTCAGCGGCACTTGCCCTTTTAGGGGTGGCCGTGAGCTATGAGCAACAGCTGTCTGA	4524
QY	2238	gaatccactaccatgagtgcgcgaatgcgaatcttactac-----cttcgatalctc	2287
Db	4525	GCCACCTGGCGCTTGCTGTCTGCTGTGACCACTTTGAGACCGCGGCCCTGCGCATGTGAACAC	4584
QY	2288	cggggaggtgcggaagtgtcatlltcaactcccaatgcccactgaagagaag-----	2340
Db	4585	CTGCAGCATCTGTGGGTGTGAGGCCACCCCTGTCTGAGGGGTACAGAGGAGGGAGGCC	4644
QY	2341	-----ccaaaggcggt	2352
Db	4645	TGAGGCGCTGCGGCGCTTCTACCCGAAATTTCTGAGACTCCCTCCGCTGCACCTTTTGG	4704
QY	2353	aagaagaagtgtatctacaacccggttactcttcaaggagaagc-----gttgaag	2402
Db	4705	ATTACGACGAGCTGTGGGTGTCCACCCAGGCGCTTTGGGTAGGCGCCCAAGGCGTGGAGAGG	4764
QY	2403	ctgcgcgaacctgtgcacggtgtgtatccaaccccccaagtggtgcatagttactcat	2462
Db	4765	CTTGCCACCGCAATTGTCTACCCACCACTGTGAAGAGCCACCACTCTCTCCCTGTGCATATAG	4824
QY	2463	gcagagactgtcaaggcctgcgccttggaagaccagatacacgtgttaacaacccggggatgtg	2522
Db	4825	CAGTGAAGTGCACAGCTTGCCCTTGCGCGGCCCGACAGACCTTGTAGTACCGTGGCGGTGG	4884
QY	2523	ccggaatcttatacacccatgggaagtgctatgtccacacgggtcttcaacgaggaagc	2582

Db	4885	CAAGAGGCGATAGTGGGAGTGGGCAAGTGTGTGTCGCTTAAAGTTTGGTGGGACAGC	4944
Oy	2583	ctcgagagctctctctgagcaatggagaatttgggcctgaactcaagcccgagctgtccga	2642
Db	4945	CTGTGAACCTTGTCTCTCTCTGTCCTTGTGGGCCCCATTTGTCAAGCTCTCCCTGTCACTGT	5004
Oy	2643	gcatggacagatgatataaggaggaatacaagctccggggagtgacctgtgtgaacaagggtg	27020
Db	5005	GCATGTGCCGCTGTATATAGGGCCTTGGGGGCTCTGGCTCTCTCTGTGTGTGATGAAGGCTG	5084
Oy	2703	gaacagcagctctgtgtgaacaatcccaagctgtatctgcagttgtgaacctgtctgtc	2762
Db	5065	GACTGGGCGCACGCTGTGTAGGTGTGCACATGTGGAGCTGTGACGCTGTGTATACCCACACCTGTGCG	51244
Oy	2763	cgtagcagccaacctgatacggagaacaagatgtgtgtgttaacttgaactcagaaagtga	28222
Db	5125	ACCCGAGGCTGTGTGCCGTGACAGCAACACTGTGAGTCAACCTGTGGCTATGAAGGGGA	5184
Oy	2823	cgagatcaatcatgcaatctgtgtgactctgtgcaacaagaacaagggggtgtgtgcgaagt	28822
Db	5185	TGGCGGTGTGTGTACACTGTGCACACCTGTGTGCCGAGCGGGCATGTGTGCTGTGACGTAGACGA	52444
Oy	2883	cgctcaagtgctccccaagaaggccccaagtctctctgtcaacttcaagaagaaggtctacaagg	29424
Db	5245	CGCAACAGTATGCCAGGTATGAAACAATGTGTCACTTGTACTCTGCTGCCGATATGAGAGGG	5304
Oy	2943	ggaatgtactacagctgtcaatagaaatagaccctgtgcagaagggtgtcaagggggtgtgtca	30020
Db	5305	TGATGTGCTGAGAGCTGTCCGGGCGCGCAACCCCTGTACACATATGGCGCACCGGGGGGCTGTGAG	5364
Oy	3003	tgaagcaagccactctcgaagatgaagggcccaaggcaagatgaatgttgatgttaaaatgtca	3062
Db	5365	CGAGACACCCCAACGTGTGTGAGCAACCGGCTGTGAACACACGCGCTGTGATGTGCACAGCG	54244
Oy	3063	ctatgtcggggagcaggttgagactgt---gagccctgaagcagctgtccgctctgaagctgtgtc	31119
Db	5425	CTACGTATAGCGATATGACTGTGCAGTGTCTGAGAGAGTGTGGAAACCACTGTGTGAGACCGTGTCT	54844
Oy	3120	acaggaacaaagagactgtgcaaccccaatgcccagctgtgtcagaacctctacttcacagagac	31719
Db	5485	GGGCGCAGCACCGCTCTGGCCCACTACATGTGCATGTGCATCTGCATCTGCATTTCCAGAGAA	55444
Oy	3180	gacagttaggtatctcaatctlaagctccccaactggccagatgaacaacagatctgttgaaca	32339
Db	5545	ACGGGCTGTGGCTTTTCCACTCTCCAGGCGCACACAGGCGCTTATGTGTCAACTTTTTCGGA	5604
Oy	3240	agccaaagaagcctgtgtccaaagaagctgtgacacatagctacactacaacagactctccta	32939
Db	5605	GGCTTAGCGCGCATGTGCACACAGGAGCGCGCTCTTCTTATTTCCCTACACTCTCTGTGC	56644
Oy	3300	tggccaagaagccaagatataacctgtgtctgcgcgcgtgtgtctgtggagatgtgggcggtgtgc	33539
Db	5665	TGCCAAGAGAGCTGGGCTTTCCACTCTGTGCTCATATGGGCTGTGGCCATATGGCTCCACTTGC	57244
Oy	3360	ctaacgcgaactcgtatgacctctcagaagatgtgtgtgcaaaagcttgttgtgaatgttgaacta	34119
Db	5725	CCACCCCTGTGTTTTTCCCTGTGGGGGAGACTGTGGCAATGTGCGGGTGGGATATGTGACGCT	57844
Oy	3420	cggatccagtgagcaacaagaagtgaatatgtgtgtgtgtctctctgttaccggatgaagaatgtc	34719
Db	5785	GGGTGTCGCCGAAGAACCCTTCAGAACCTGTGGATGTGCTACTGCTTCCGTGTCAAGAGAT	58444
Oy	3480	gaactgtgacccgtgaagcagagatatactgtgtggagatgtg---ctctctgtgaagtggaaact	35386
Db	5845	GGCTTCCGATGCCGAATGTGCTTCTGTGGGTGACGGGATTCAGCACTGTGCATATGGGAAGCT	5904
Oy	3537	gtctgaagttctctatgtctcttccctctgcgtcaaaaactctctgaacagaggtgtgtgttct	3596
Db	5905	GCTGTATGTGTGTGCTGTGCACACTGTGCACACTTCTCTACCTTTCTATGGGATGTATTTGGGCTGA	5964
Oy	3597	ttcaagaagctcagcccgagagacaggaatttttgaacaacctgtgactgtacactgttccatccg	3656
Db	5965	TGCCATATCCACCCAGCGGGGTCTGTGCACTTCTGTGAGACTTCTGTGATGATGAGACTACGTA	6024

[illegible]

RESULT	11				
LOCUS	HSA275213	7870 bp	mRNA	linear	PRI 29-MAR-2000
DEFINITION	Homo sapiens mRNA for stabilin-1 (stabl gene).				
ACCESSION	AJ275213				
VERSION	AJ275213.1	GI:6469373			
KEYWORDS	stabl gene; stabilin-1.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
AUTHORS	1 (bases 1 to 7870) Pollitz,O., Guillot,P., Gratchev,A., Schledewski,K., Birk,R., Hakly,N., Tebbe,B., Orfanos,C.E. and Goerd,S.				
TITLE	Stabilin-1, an endothelial-macrophage member of the fasciclin domain containing protein family associated with angiogenesis unpublished				
JOURNAL	2 (bases 1 to 7870) Pollitz,O.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (25-NOV-1999) Pollitz O., Department of Dermatology, Benjamin Franklin Medical Centre Free University Berlin, Hinderburgdamm 30, Berlin 12200, FRG				
JOURNAL	Related sequence D87433.				
COMMENT	Location/Qualifiers				
FEATURES	1..7870				
Source	/organism="Homo sapiens"				
gene	/db_xref="taxon:9606"				
CDS	31..7743				
	/gene="stabl"				
	31..7743				
	/gene="stabilin-1"				
	/codon_start=1				
	/product="stabilin-1"				
	/protein_id="CAB61827.1"				
	/db_xref="GI:6469374"				
	/translation="MAGPRGLPLCLLAFCLAGSPFVQGVLFKCDVKTFTVTHVHC				
	TSCAIKKOTCPGSMRLPOLIITODRYEVLQGSYMSGCKRKORQVYKAAACP				
	GYMSRCHCGGAGETPCNGHGTLDMDRNGTVCQENPGSKCCOCPDNRRGPPQ				
	OSVSCVHGYNBHPRGDCSLCAVGTGHRGDELEFCOELRPQNTQCAEAPSCR				
	CLPRTDGGSECRAPNCPWSPKSLQCVSPFKGQAOCHCPENYHDDGMVCIKPDCC				
	TIDNGGSENSTLIVOKPGAGFCTGRGLVSIINSAAGCFACSPSCSRKTCQV				
	TADKTSVCSESESDGGRACVGHLLHEVQKATGTGKVFQIQLRVAVMMDGCKREILT				

Db	3784	gmcctgacgctggagcgtcaagtgccgttcggatattaccacacgcttgagagcccttcgcggagagaaa	3843	tacgpttvlvpsvssfssttmasslaqlcrrhliiaaghlledrtootrramwllaog ltvtfnoftstiyktykdpoofntvityanmaiaagvthvtglmwqpsvsgpdkprt igolltaftearspftvlennglpsliidndppfvtvpsnvaadldsgrltylftgag lskalelvaryhiynhgolpvtelkskrlitmanovylansegrllilepvylovr vdmaangvianhmlldlptlplilpkhsseepohkiavagcvcvolantstcpnsln klclpkeevcyihdpctlnvlykkccasyctntimeocckgfpedctocrgsnoc ykcncsdgdlognkcacilcpdykktlacthsnphkheociedcvgchldnkrpgsg cvcoggtcdaepsgsrpfcnesmgdcgpgllaothclnarvcysgebvavarkclidcgeggoo fscprnscshpdrdgscsenaevcagsgtghcchcagwsdgdcgvcaidecylavgsrclh gchfdalcslvygpcscroctcklfgpavgoyocspidprcagmgcghcgvatiravdgogor vctprpbgdgsfscvgdipreleannhfslyfomksagittlpadrbarylaseaa vrojlspebrapwldprtlpnlvyrhpflogalarlbgoevatllnpttrmetrni sgrwvovasyadvadllatngvlihlisovlllprgdvpgogllqolldlvapslpre lllohhglvpolleatvattienvtrnrsleaoqnsnshldadryrhnhvigealsmetlkrk gghnshlspoliahmivlyvfnhsogepvnhpllegmlenaprslielscylvgsrrclh shabarckxvncvtrrfrcrtogfopolodprksyechgscfscstakrktiovpnc crgpftlclercpbgclgycsgchqocdrplrgsgechhcfhrtacveclrgpnc ctygcddahgicldqcllogdsgcvcnvmogrlcookiitpocprkcpnannvcodsag astcaagaysgngifpcesvdpnpgogscspkncvkaivpgocttccogymwgdsl coeinsclihhgghchihaeciprcpooqscsgrpvcgsgdiirctelldpcsknngcs pyatckstgdcortctcdattvagdltlrcarvlelllrokhasfelfrlleykelg dpattfvyphadlmslnsodelarlrarholrvfyrhvgorrlsefllrogyatals ghlrlrsresbsiylndfparvssdheavngtlhfldrvllprealhwepddapirpr nvtaaogefgtkifsgllkvaagllplrlenshrfptlmpidrafrpelpdromwlyh edhdhklalailrgmhrnvevaladlplnlgrlrhmgtpfplsfsrarpelpdromwledb ariivorhlpregclavgldolleprrgcargdheetrlprlnticslglprrpcesso eogsepacmrfyepkfmtyplshlslglsrswahvplmwrpoglrgchgnrnlvttmwrps ccphgysceogacggpspscdprgocmwcmsgocslscsfagpctcelcaraprp hcoqrcrtvghrcdcbglgsgscscrcdrgmwtprcrevlellopvctppcarbaycraa scbecslsebdgrvctvadiicdohdgscshpansovgtwctclpdyrbgdsmsraa rnpctdghrgcshpamcilstgltrrcbcehagvsgldoclesepvrpcolgopp chspamcldhofekragcvfltoatngsrpafnfeaaeacvaylaspolsaaoo lgfhlclmglmangstnaphvypfvadcgngvstafygarlnnisermdaefrvroday crgnrgfvgdgiistcnkllldvlatnatspnycmllgynanarogldpfdldelt yktlrfvynbsfvdnmtlsgpdelhlasmnatlllsanaasogkllpnhsglslitlsdap dnsswapaepgtvavsrliymwlmfmgfllhalspllapqpavylaprapvyaav gavlaaakallglvagalylrarngkprtcfgtsafaedaddddespmoeglnptlvsvp npvgsdipfcppddslleedpftolrlvfk"
BASE COUNT 1420 a 2482 c 2446 g 1522 t				
ORIGIN				
Query Match 14.8%; Score 697.2; DB 9; Length 7870;				
Best Local Similarity 51.5%; Pred. No. 6.9e-188;				
Matches 2066; Conservative 0; Mismatches 1823; Indels 125; Gaps 15;				
QY	25	ctcgagcagaatgcccgacatccatttccagagattacattatccataaaccctgca	84	
Db	3433	ctgagcttggtgccccttgcagcccttccgggaattgtgcagaccacccattgggttg	3492	
QY	85	agtgaatccagatctgcagatgcttaactatgttctgcgtgcaaaataagaagccat	144	
Db	3493	ccccgatttgagggctgcacaccccttaccacccatttttggtccacccacacccgt	3552	
QY	145	aactatccaggaagaagaagccacatcctctaaaggaataatctcagtcagcatg	204	
Db	3553	gcc-----cagggacacacgacgtaacctgcagacgacgacacacagtgccgac	3603	
QY	205	gtcctgagggaaagaagctcctctgaagaatgactgcataaagcgatgcactcgagag	264	
Db	3604	gmcctgaggagagcccttccatggaacccctggagaggggagacacccgacaaccc	3663	
QY	265	ctggaggtctcttaactccttgctcttcttctccgaaatgaccagctatgtaaa	324	
Db	3664	ctgggcccctgcccattgcatgcttcttcaacaaccactatggccacacccagctga	3723	
QY	325	gctccaataaactacacaaatgtygcaactgataaagagtgcattcatgctgagaa	384	
Db	3724	gtggccactgggaagccccaatgcttggaagcccttgagcccgctgcgtgatttg	3783	
QY	385	gtcttggaatctcagaagaacagatgtgacaaataatgacacacattatgttga	444	
Db	3784	gmcctgacgctggagcgtcaagtgccgttcggatattaccacacgcttgagagcc	3843	



QY	445	tgtygaagtgltccaaagaagcccccctgcgacactcgagacaaaaccacttagagagacg	504
Db	3844	TTTTTAACTGCACAGAGGATTTCCGCTGCACCTACAGGCGCTTCCAGCTGCAGACACACCC	3903
QY	505	aggaatgcatctatctcaactctactctcaatgggaagatccggtatctacgtgggtccag	564
Db	3904	AGGAAGAGCTGTGTCTACCGATCTGGCTT-----CTCCTTCCCGGGGCTCTCT	3954
QY	565	ccacagtgtytagaacaacatctacaagagcctgty-gctggtctcttggccacaa	623
Db	3955	TTACACATGTGCCAAGAAGATTCAGGTGCCGGAAGCTGCTGCCCTGGTTCTTTGGCAGCGTG	4014
QY	624	tgccaagcctgcgcccgagaggtccaatgtgtgtctctgtygaagcgcttctgtctgga	683
Db	4015	TGTGAGCCATGCGCAGGGGGGCTTAGGGGGGTTGTGCTCAGGCCATGCGGAGTGCC-AGGA	4073
QY	684	cggtgtgaatgagcactgagacgttgcacagtgcgggtgcgtgcgttcaatggagacgcgtga	743
Db	4074	CAGGTTCTCGGGACAGGGGAGTGGCACTGCCACGAGGGGCTTCATATGAAAGCGCCTGTGA	4133
QY	744	aaccgcaactgaggagaaatgatgatctccacttgcacaaagacatgtctctgttccatgt	803
Db	4134	GCTGTGTAGGTGGGGCCGCTACGGGGCCCAACTGCACCGGAGTGTGATGATGCCCATGG	4193
QY	804	gagatgtagccaaggaaccccttgggagaagcgctccctgtgaactgtgaagtcgctggcgag	863
Db	4194	GCTGTGCCAGGAGGGGCTCAAGGGGACGGAAGCTGTGTATGAACTGTGGCTGGGAGGG	4253
QY	864	agtgaagtgtgacatlgagatcacacacagacactgacaagagactgttcaacacagtc	923
Db	4254	CCTCGGCTGTGACAGAAATATCACCGCCCTCAGTGCCTTAGSAAATGCGAGCCCCATGC	4313
QY	924	caactgcctctgtgalccagacggaagcctcgtycaaatgtlgtgggaagatccgaag	983
Db	4314	CACATGCGTGACAGACTCCGGCGGAGCTTCACACTGCGCTGTGCTGGCGGATATCTCCG	4373
QY	984	gaatgaagcgtgtgcagacacatcaatgtccttgaagcagaatggaagtgtcttac	1043
Db	4374	CATATGCATCTTCTGTTTCAGAGGTGAGCCCTGCGCCACAGGCCATGGGGGCTGCTCCC	4433
QY	1044	aaagcgcgactgtaaaagaacacccacagaaaccgggtgtgtgtgtgtgaagaagcaagcta	1103
Db	4434	TCATGCCAATCTATACCAAGGTGGACCTGGGACGCGAGATGCACCTGCCAGATGGCTA	4493
QY	1104	taccggcgacggcatcgtgtgtccttgtaaatcaaccgggtgtttggagaaccaaavgtgctg	1163
Db	4494	CATGGCGCGAGGGAGCTGTGCCAGGAAATTAAACAGCTGTCTATCCACACCGGGGCTG	4553
QY	1164	tgacgaatgtgagagtgacacagacagggcccaacagggccgtgtgtgaagctgtgc	1223
Db	4554	CCACATTCACGCGAGTGCATCCCACTG6CCCCCAGCAGAGTTCCTGCGAGCTGCCGTGA	4613
QY	1224	gaagtacacttggatlgaa---aggltctgtcgtcttcaatgtlctgtgcclaagaca	1280
Db	4614	GGTTTACAGCGGGAGTGGCATCCGAGCTCGGAGCTTCGTGGACCCCTGCTTAAGACAA	4673
QY	1281	tgcgagctgcagtcacattgtccttctgcacactacactgaagcaagatcaagatagtac	1340
Db	4674	TGAGAGATCGAGCCCATATGTCACCTGCACAAACACAGGGGATGGCCAGAGGACATGTAC	4733
QY	1341	ctgcaagcagactaac---gggtgatggaatcgtctgcggggagcaactcaacgggga	1397
Db	4734	CTGCGACACAGCCACCGTGGGGGACCGCTTACCTGCGGTCGGAGCTGCGCTGGGA	4793
QY	1398	gcttcccaagaaccccttcgacgtccccagtaactcttccagttgcaggaagatgctgcg	1457
Db	4794	GCTCCTGAGAGGAATAACCATGCTCATTTCTTACGCTCCCGCTCCTGSAATAT-----AA	4847
QY	1458	agagcttgctgaacgtgcgcccttcaacgltgtcgccgcttvtctagctcttcaatca	1517
Db	4848	GGAGCTCAAGGGCGATGGGCTTTTACACATCTTCTGTCGCCACGACAGATTAATGAGCAA	4907



D	5985	CATGACGGCATAGTGGCAGTGGGCAAGTGTCTGTGCCCTTACGCTTTTTCGTGGGACAGC	6044
Q	2583	ctgagcagctctgctgcatlbgagatlltggctctgactlcaagcccgagctgctccga	2642
D	6045	CTGTGACACTGTGTGCTCTCTGTGTCCCTTTGGGCCCATTTGTCAAGCTCGCGCTGACTGT	6104
Q	2643	gcatlbgacagltgatlbgatggggaatcaacagctccggggagtgctctgtlgaacaaggtg	2702
D	6105	GCAATGCGCGCTGTGATGATAGAGGCGCTTGGGGGCTCTGTGCTCTCTGTGATGAGCGCTG	6164
Q	2703	gaagcgcgctctgtgtgagcaactcccaagctgtatctcgagtgtygaacactgtgtctc	2762
D	6165	GACTGGGCCACGCGTGTGAGGTGCACACTGAGCTGCAGCGCTGTGTATACCCACCCCTTGC	6224
Q	2763	cgltgcagccagccctgtlaacgagaacagacgtgtgtgtltaactlgaactaagaagtgga	2822
D	6225	ACCCGAGAGCTGTGTGCCGTGCAGGCAACACTGTGTAGTCTACCCCTGGCTATTGAAGGGGA	6284
Q	2823	cgagatcacatlgacagctgtgtgactcttgcaacaacgaacaaacgggggctgtgcgaagt	2882
D	6285	TGGCGCGGTGTGATAGTGTGCACACTGTGTCCAGGACGGGATGGTGGCTGATGAGCA	6344
Q	2883	cgctaaagtgtctccagaagaagcgcccaagtctctctgtcagctgtcaagaagaagctacaagg	2942
D	6345	CGCCAACTGTGAGCAGGATAGGAACAATGTGCATTTGTACTCTCTCCGCGGACTACGAGCG	6404
Q	2943	ggaatgtctacagctgtgcaatagagatagaccctgtgacacggtytcaacgggggagatgca	3002
D	6405	TGATGTCTGAGAGCTCGCGGGCCGCGCAACCCCTGTGCACAGATGGCGCACCGCGGGGCTTGAG	6464
Q	3003	tgagcacgcacactctgcagagaatgaacggggccaggaacgaatagltgaaatgaagaatga	3062
D	6465	CGACGACCGCCAACTGCTTGTAGCACACGGGCGCTGAACACACGGGCGCTGTGATGCCACAGG	6524
Q	3063	ctatgtctggggagcggagatgtgactgtgagcttgagcaagtgcgcg---tcagcgttgtct	3119
D	6525	CTACGTAGGGGATGTGACATGTGCAGTGTGTGAGGAGTGTGGAACCACTGTGTGACACCGCTGCTT	6584
Q	3120	acagagaacagagacagatggccaccagatgttcaactcttgaagaccctacttccaaagac	3179
D	6585	GGCGACCAACCGCGCTCGCCACTCTCAATGCATGTGCATGTGCACAGCACTTCCAGGACAA	6644
Q	3180	gaccgttagagatltccatctacatgcctcccaactgtggcgacatgaacaaactgtacatltgaaca	3239
D	6645	ACGGGCTGGGGTTTTCACACTCCAGGCGCAACAGCGGCGCTTATGGTGTGAAGCTTTTCGA	6704
Q	3240	agccaaagaagcctgtgtgccaagaagctgtgacatagccacatcaaaccaagctctcta	3299
D	6705	GGCTGACGGCGGATGTGCACAGCACAGGAGACCGCTTCTTATTTCCCTCACACTCTTCGC	6764
Q	3300	tgcccagaagaagcaagatataactctgtgtctcggcgccgtgtgtctgtgagagatggggctgc	3359
D	6765	TGCCCACTAGCTGGGCTTCCACTCTGTGCCATGTGSGCTTGCTGCCAATGGCTCCACTGC	6824
Q	3360	ctaccgactacatgtgtcctctcagaagatgtgtgtgaacaaagtltgtlbggactgtlgaacta	3419
D	6825	CCACCGTGTGTGTTTCCCTGTGTGGCGGACGTGTGCATGTGTCGGTGGGTAGTGAAGCT	6884
Q	3420	cgatctcagagccacaacaagaatgtgaatgtggagatgtctctgttccgagatgaagaatgt	3479
D	6885	GGGTGCCCGGCAAGAACCTCTCAGAACGCTGTGGGATGCTCACTCTTCCGTGTCAAGATGT	6944
Q	3480	gaactgtacactgtcaaggcaagatgtlbgagatgg---cttctgtlbgagtggaacct	3536
D	6945	GGCTCGCCATGCGGAATGGCTTTGTGTGGTGCAGGGATTCACACACTGTCAATGGGAACCT	7004
Q	3537	gtctcagatctctcatgtcctctccctctgcgtcacaacactctccagacagagtgctgtctc	3596
D	7005	GCTGGATGTGTGTGCTCCCACTGCCAACTTCTCCACTTCTATGGGATGTCTATTGGCTA	7064
Q	3597	ttccaaagatctcagcccgagagacagcatlltgaacaaactgtacactgtlccatcgcg	3656

Db	7065	TCGCATATGCCACCCACAGCGGGGCTCTGCAGCTTCTCTGCAGCTTCTCTGGATGATGAGCTCACGTA	7124
Qy	3657	tggaacaccggtttgttgcacacgaacagtggtgttgcacggggaataaagagcgttctgtgcgc	3716
Db	7125	TAAAGACACTCTTGTGTCGTCGTAATGAAAGGCTTTGTGGACAACATGAGCCTGATGGGCC	7184
Qy	3717	ggacattgagacacacactcaatcatgtcaaggtctctcttcaalgtac-----ctt	3768
Db	7185	AGACTTGGAGCTGATGAGCTTCCACAGCGCACCTCCCTTAATGTGCACAGCGCACGAGGAA	7244
Qy	3769	gtcaatgtgaacttcttgaagactatgtcgggaagcaactgtctattcccttgaagcag	3828
Db	7245	GTTCGTCCTCGGCGCACCTCGAGCTTCAGCTTCATCATGATGAGCGAGGCGTGGACACG	7304
Qy	3829	gaccagctccacacgaagacca-aggtttggatbgaagatccatctgtcagtggaat	3887
Db	7305	TTTCCTGGGCGCCCTGTGCGCCCAAGGACAGTGTGTGTTAGCGGTATCATTTGTGTGGACAT	7364
Qy	3888	catgcgcgcgaatbgaatccctccatattattcttgaaccttgaagctctcc	3941
Db	7365	CATGGCTTCATGATGAGCATCATGATCATGCTGTGCGCACGCCCTCTGCGCACCC	7418
RESULT 12			
AX149457			
LOCUS	AX149457	1624 bp	DNA
DEFINITION	Sequence 13 from Patent WO0136638.		
ACCESSION	AX149457		
VERSION	AX149457.1	GI:14347934	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 1624)		
AUTHORS	Shimkels,R.A., Lichenstein,H., Vernet,C. and Fernandes,E.		
TITLE	Polypeptides and nucleic acids encoding same		
JOURNAL	Patent: WO 0136638-A 13 25-MAY-2001; Curagen Corporation (US)		
FEATURES	Location/Qualifiers		
source	1..1624		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	417 a 420 c 386 g 401 t		
ORIGIN			
Query Match	13.5%;	Score 637.6;	DB 6;
Best Local Similarity	76.0%;	Pred. No. 7.2e-171;	Length 1624;
Matches 830;	Conservative 0;	Mismatches 224;	Indels 38;
		Gaps 2;	
Qy	1508	ccctcaatcatbagaagcccgatlaaagactbgaagcagcagggccctcatgtccaggttc	1567
Db	571	CCCTGATCTTCTGTATTCAGGTTAAAGATGGGACAAATACGGTTAAATGCCAGGTTTC	630
Qy	1568	tctgcgtatcacgttggtggtgtccagcagctgtctgttggacacactaaagtgtcacaca	1627
Db	631	TTTCGTACCATTTGTGTGCGCTGCCACACACTCTTCTGTGAAAAACCTGAATTGATCTCA	690
Qy	1628	gttcacagacccctccaagagagacaggttccatctgtcttctcagaagactgtgtca	1687
Db	691	ATGCTACTTCCCTCCCAAGAGAGACCAATAGTATCTCCCTCTCTCAGACACGGTGTATA	750
Qy	1688	taaaacatbagaagcgaaggtcgtctcagtgacatataagcaccgaatgtcgtcaluccag	1747
Db	751	TAAACAAATTAAGGCTTAAGATCATATTCAGATATCATCTAGTAAATGGGATTTTCATA	810
Qy	1748	ttatagacacagttgtcgtcttcccaaaacttcttatacaccaccaaga tgccttggga	1807
Db	811	TCATGACAAATTTGCTATCTCCCAAAAATTTGCTTATCTACCTCCAAAGCAACTCTGGAA	870
Qy	1808	gggttctgtcaaaactcttactacagttgtgcagcaaacaccagatatatacacaattcagcaat	1867

Db 871 GAATTCGCAAAATCTTACGACTTTGGCAACAAACAAATGGCTACATCAATTAATGCAACT 930  
 QY 1868 tgatagaagactaaggtctgctcaatcaactgaactcatcacacccagctcag 1927  
 Db 931 TAATACAGGACTAGGTTTCTGAGTGTCAACCGATCCCATCCACACCCAGTCACTC 990  
 QY 1928 tctcttgctcagacagacagccctggaagccttgccccagagacagagactcctgt 1987  
 Db 991 TCTTGTGGCCACGACCAAGCCCTCCATGCCCTACCTGCTGAACAACAGACACTTCTGT 1050  
 QY 1988 tcaatcaagacacaaagacaaagctgaagcttcaactgaagttccagttatccgaagact 2047  
 Db 1051 TCAACCAACCAACAAAGGACAAAGCTGAAGAGATTTGATTTGATGTATACGAGATG 1110  
 QY 2048 ccaagagtttagctcagacccctccaggtcctgtcttccctggaagacccctgaagctcag 2107  
 Db 1111 CCAAGGTTTAACTGTGTGATCTTCCACATCCACTGCTGTGAACACCCCTGCAAGTTTCAG 1170  
 QY 2108 agctgaagtgtgaaggtgtggaacttgacatcagtgatcgttcttctaaacgaacaaa 2167  
 Db 1171 AGCTGAGTGTGAATGTGAGACTGGACGAGCATCGGATCGGACTCTTCTGAAATGGCCAAA 1230  
 QY 2168 tctgcaaatcaatcaacacccggagactctgtttgacgtgtgtgtgctatgagcttga 2227  
 Db 1231 CCTGCAAAATGTGTCAAGCGGAGCTCTTGTGACCTGGGTGTGGCTTACGCGCAATGTACT 1290  
 QY 2228 gctcactcaatgaactcactcaactgaagtgacatgagcttactcaacttccatc 2287  
 Db 1291 GTCTGCTGATGTATGCCACCTGGGGGGCCGCTGTGACACCTTTACTTACTTCCATGCTCT 1350  
 QY 2288 cggggaggtgaggaagtgtcatcttccatcccaaatgcccacatgaagaagaaccaaag 2347  
 Db 1351 CGGGGAGGTGTGGAGGCTGTGTCAATCTCCAGCTGCCCAAGGGAGGATTAACCAAAAGG 1410  
 QY 2348 gctgtgaagaagaagtgtatctcaacacccgttacccttcaaggaagaagtggaagctgccc 2407  
 Db 1411 GTTGAACCAAGATGTCTTACCAAC--TGCCCTTTAAAGAGAACTGTGAAGGCTGCC 1467  
 QY 2408 agaacccttgacacgtgtgtatccaaaccccaagtgctgcatgttacttcatgacag 2467  
 Db 1468 GGGAGCGGTGCAAGCTGTGATACGATCCCA----- 1500  
 QY 2468 actgcaagcctgacctgaggaagacagataacacgtgtaaacacccggagcatgtgcag 2527  
 Db 1501 -----GCTGCCCCCTGAGAGACCAAGATGCCCTGTAAATAACCGGGGTGTGCTCTTG 1552  
 QY 2528 atcgtatacaacacatggaacagtgctatgccaacccgcttcaacggaggaagcctgcg 2587  
 Db 1553 ATCACTACTTCCGCCACCGAGAGTGTAAATGCAACACCGGCTTCAATGGAGCGGCTGTG 1612  
 QY 2588 agctctgtctgc 2599  
 Db 1613 AGATGTCTCTGGC 1624  
 RESULT 13  
 AX149463 1577 bp DNA linear PAT 08-JUN-2001  
 LOCUS AX149463  
 DEFINITION Sequence 19 from Patent WO0136638.  
 ACCESSION AX149463  
 VERSION AX149463.1 GI:14347937  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 1577)  
 AUTHORS Shimkets,R.A., Lichenstein,H., Vernet,C. and Fernandes,E.  
 TITLE Polypeptides and nucleic acids encoding same  
 JOURNAL Patent: WO 0136638-A 19 25-MAY-2001;  
 Curagen Corporation (US)  
 FEATURES  
 SOURCE 1..1577

BASE COUNT 424 a 410 c 342 g 401 t  
 ORIGIN  
 Query Match 11.0%; Score 517.4; DB 6; Length 1577;  
 Best Local Similarity 78.8%; Pred. No. 2e-136;  
 Matches 617; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 QY 1508 ccttcaatcagccccgagatlaaagacttggatcagacagagcctcatgtlcccaagttc 1567  
 Db 571 CCTGTGATCTTCTGTACAGGTTAAACAGCTGGACAAATACGGTTTATGGCCAGGTTTC 630  
 QY 1568 ttcgatacaagtggtgtgctgcagacagctgtctgttggaacactaaagtgaccacaa 1627  
 Db 631 TTCGGTACATGTGTGTGCTGCCACCAAGCTGCTTGTGAAACCTGAAATGTCTCA 690  
 QY 1628 gtgcacagacccctcaagggagccagtttccatctctgtctcagaagactgtttca 1687  
 Db 691 ATGCTACTTCCCTCAAGGAGAGCAATATGATCTCGTCTCAGAGCAGGTTGATA 750  
 QY 1688 taaacaatgaagcgaaaggtccctgcagtgacatcatcaagaccaatgtgtcalcagc 1747  
 Db 751 TAAACATTAAGGCTAAGATCATATCCAGTGTATCATGATCAATGAATGGATGTTCATA 810  
 QY 1748 ttatagaagaatgtctgtctcccaaaaactgtctatcaaccccaagaatgtctggca 1807  
 Db 811 TCATAGACAAATGTCTATCTCCCAAAATTTGTATCATCTCCCAAAACCAACCTGTGAA 870  
 QY 1808 ggtgtctgcaaaacttactcagtgtagcaaaacccaagagatatataccaatcagcaagt 1867  
 Db 871 GAATTCGCAAAATCTTACGACTTTGGCAACAAACAAATGGCTACATCAATTTAGCAACT 930  
 QY 1868 tgataagaagactcaggtctgtgtcagatcatcactgaacttccatcacacccagctcag 1927  
 Db 931 TAATACAGGACTCAAGTTTGTGAGTGTCAACCGATCCCATCCACACCCAGTCACTC 990  
 QY 1928 tctcttgctcagacagacagccctggaagccttgccccagagacagagactcctgt 1987  
 Db 991 TCTTGTGGCCACGACCAAGCCCTCCATGCCCTACCTGCTGAACCAACAGACACTTCTGT 1050  
 QY 1988 tcaatcaagacacaaagacaaagctgaagcttactcaactgaagttccagttgacagact 2047  
 Db 1051 TCAACCAACCAACAAAGGACAAAGCTGAAGAGTATTTAAGTTTATGTGATACGAGATG 1110  
 QY 2048 ccaagcctttagctcagacacccctcgaagctgtcttccctggaagacccctgcaagctcag 2107  
 Db 1111 CCAAGGTTTAACTGTGTGATCTTCCACATCCACTGCTGTGAAGGCTGCCAAGGTTTCAG 1170  
 QY 2108 agctgaagtgtgaaggtgtggaacttgacatcagtgatcgttcttctaaacgaacaaa 2167  
 Db 1171 AGCTGAGTGTGAATGTGAGAGTGTGAGGACATGGGTGACCTCTTGTGAATGGCCAAA 1230  
 QY 2168 tctgcaaatcaatcaacacccggagactctgtttgacgtgtgtgtgctatgagcttga 2227  
 Db 1231 CCGTGAAGATTTGTGACGGGAGCTTGTGTGACTGGGTGTGACTTATGGCATTTGACT 1290  
 QY 2228 gctactcaatgaactcactcaactgaagtgagcgaatgtgacacttactcaactcgaatlc 2287  
 Db 1291 GTCTGCTGATGTATCCCAACCCGCGGGGGCGCTGTGACACCTTTACTACTTGCATGCT 1350  
 QY 2288 cgg 2290  
 Db 1351 CGG 1353  
 RESULT 14  
 AC025501 218073 bp DNA linear HTG 25-JUN-2001  
 LOCUS AC025501  
 DEFINITION Mus musculus chromosome 10 clone RP23-129023 map 10, WORKING DRAFT  
 SEQUENCE, 18 unordered pieces.  
 ACCESSION AC025501

VERSION AC025501.4 GI:14547868  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 218073)  
REFERENCE Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Mus musculus chromosome 10, clone RP23-129023  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 218073)  
REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,  
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, D.S.,  
Dodgson, S., Domingo, M., Doyle, M., Ferrelle, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczek, J.,  
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R.,  
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Teste, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, R., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (09 MAR 2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 25, 2001 this sequence version replaced g1:8077096.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L8063  
Center clone name: 129\_O\_23  
----- Summary Statistics  
Sequencing vector: M13; M7815; 44% of reads  
Sequencing vector: Plasmid; n/a; 56% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 213014 bases at least Q40  
Consensus quality: 214764 bases at least Q30  
Consensus quality: 215528 bases at least Q20  
Insert size: 210000; agarose-fp  
Insert size: 216373; sum-of-coverage  
Quality coverage: 10.2 in Q20 bases; agarose-fp  
Quality coverage: 9.9 in Q20.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 23321: contig of 23321 bp in length  
\* 23322 23421: gap of 100 bp  
\* 23422 24529: contig of 1108 bp in length  
\* 24530 24629: gap of 100 bp  
\* 24630 26959: contig of 2330 bp in length  
\* 26960 27059: gap of 100 bp

FEATURES  
source  
\* 27060 28240: contig of 1181 bp in length  
\* 28241 28340: gap of 100 bp  
\* 28341 30420: contig of 2080 bp in length  
\* 30421 30520: gap of 100 bp  
\* 30521 32131: contig of 1611 bp in length  
\* 32132 32231: gap of 100 bp  
\* 32232 34170: contig of 1939 bp in length  
\* 34171 34270: gap of 100 bp  
\* 34271 36716: contig of 2446 bp in length  
\* 36717 36816: gap of 100 bp  
\* 36817 40349: contig of 3533 bp in length  
\* 40350 40449: gap of 100 bp  
\* 40450 43365: contig of 2916 bp in length  
\* 43366 43465: gap of 100 bp  
\* 43466 46874: contig of 3409 bp in length  
\* 46875 46974: gap of 100 bp  
\* 46975 51691: contig of 4717 bp in length  
\* 51692 51791: gap of 100 bp  
\* 51792 118666: contig of 66875 bp in length  
\* 118667 118766: gap of 100 bp  
\* 118767 133750: contig of 14984 bp in length  
\* 133751 133850: gap of 100 bp  
\* 133851 151236: contig of 17366 bp in length  
\* 151237 151336: gap of 100 bp  
\* 151337 174475: contig of 23139 bp in length  
\* 174476 174575: gap of 100 bp  
\* 174576 208697: contig of 34122 bp in length  
\* 208698 208797: gap of 100 bp  
\* 208798 218073: contig of 9276 bp in length.  
Location/Qualifiers  
1. .218073  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="10"  
/map="10"  
/clone="RP23-129023"  
/clone\_1bp="RPC1-23 Female Mouse BAC"  
1. .23321  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
23422. .24529  
/note="assembly\_fragment"  
24630. .26959  
/note="assembly\_fragment"  
27060. .28240  
/note="assembly\_fragment"  
28341. .30420  
/note="assembly\_fragment"  
30521. .32131  
/note="assembly\_fragment"  
32232. .34170  
/note="assembly\_fragment"  
34271. .36716  
/note="assembly\_fragment"  
36817. .40349  
/note="assembly\_fragment"  
40450. .43365  
/note="assembly\_fragment"  
43466. .46874  
/note="assembly\_fragment"  
46975. .51691  
/note="assembly\_fragment"  
51792. .118666  
/note="assembly\_fragment"  
118767. .133750  
/note="assembly\_fragment"  
133851. .151236  
/note="assembly\_fragment"  
151337. .174475  
/note="assembly\_fragment"  
174576. .208697  
/note="assembly\_fragment"



```
*      87869      91230: contig of 3362 bp in length
*      91231      91330: gap of unknown length
*      91331      95333: contig of 4003 bp in length
*      95334      95434: gap of unknown length
*      10082      10081: contig of 4648 bp in length
*      10082      10081: gap of unknown length
*      100182     102921: contig of 2740 bp in length
*      102922     103021: gap of unknown length
*      103022     107381: contig of 4260 bp in length
*      107382     107381: gap of unknown length
*      107382     110228: contig of 2847 bp in length
*      110229     110328: gap of unknown length
*      110329     114370: contig of 4042 bp in length
*      114371     117842: gap of unknown length
*      117843     117942: contig of 3372 bp in length
*      117943     122460: gap of unknown length
*      122461     122560: contig of 4518 bp in length
*      122561     126540: gap of unknown length
*      126541     126740: contig of 4080 bp in length
*      126741     130890: gap of unknown length
*      130891     130990: contig of 4150 bp in length
*      130991     134751: gap of unknown length
*      134752     134851: contig of 3761 bp in length
*      134852     138027: gap of unknown length
*      138028     138127: contig of 3176 bp in length
*      138128     141177: gap of unknown length
*      141178     141278: contig of 3050 bp in length
*      141279     143823: gap of unknown length
*      143824     143923: contig of 2546 bp in length
*      143924     146573: gap of unknown length
*      146574     146673: contig of 2650 bp in length
*      146674     149832: gap of unknown length
*      149833     149832: contig of 3159 bp in length
*      149833     152476: gap of unknown length
*      152477     152570: contig of 2538 bp in length
*      152571     155089: gap of unknown length
*      155090     155189: contig of 2519 bp in length
*      155190     157216: gap of unknown length
*      157216     157216: contig of 2027 bp in length.
```

## FEATURES

```
Source          Location/Qualifiers
1..157216       /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="3"
                /clone="RP11-54J3"
```

```
BASE COUNT      44517 a 32787 c 33678 g 42931 t 3303 others
ORIGIN
```

```
Query Match      3.1%: Score 146.8: DB 2: Length 157216:
Best Local Similarity 83.8%: Pred. No. 5.1e-30:
Matches 166: Conservative 0: Mismatches 32: Indels 0: Gaps 0:

QY 2837 cagtcgtgactctgcacaacagaacaaggggctgtgcgaagctcgaagtcctcc 2896
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114169 CAGTTGTGATTTTCGAACAGCAACAGGGGGCGCTGTGCAAGAGTGCCAGATGCTCCC 114110

QY 2897 agaaagcaccacgaagtccttcgcagctgcgaagaaagctcacaaaggggagtgctacagct 2956
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114109 AGAAGGGCGACGACAGCTCTCCAGAGCTCCAGAAAGGATACAAAGGGGACGGGCACAGCT 114050

QY 2957 gcatagaaatagaccctctgcagaacagtgctcacaggggagtgcatagacagccacct 3016
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114049 GCACAGAGATGAGACCCCTGTGTCAAGAGCGCTTAACGAGGGTGTACGACAGCGCACCT 113990

QY 3017 gcagagatgacgggccag 3034
      | | ||| ||| |||
Db 113989 GTAAGATGACAGGCCCG 113972
```

```
Search completed: June 17, 2002, 16:45:43
Job time: 12758 sec
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2002, 14:00:15 ; Search time 465.74 Seconds  
(without alignments)  
17348.305 Million cell updates/sec

Title: US-09-842-930A-1  
Perfect score: 4706  
Sequence: 1 tcttaccacaagtctactcac.....aagcaaaaaaaaaaaaaa 4706

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N\_Geneseq\_032802:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4706	100.0	4706	ABA04648	Rat Hyaluronic Aci
2	2835.2	60.2	4576	ABA04662	Human Hyaluronic A
3	1599.4	34.0	3625	AAF87120	NOV9 coding sequen
4	1244.8	26.5	2483	AAF87119	NOV8 coding sequen
5	1243.2	26.4	2512	AAF87117	NOV6 coding sequen
6	1196	25.4	2011	AAF87113	NOV3 coding sequen
7	1016.4	21.6	1804	AAF87114	NOV3 coding sequen
8	883.4	18.8	1259	AAA57365	DNA encoding a hum
9	695.6	14.8	6761	AAA57362	DNA encoding a hum

10	654.4	13.9	1377	21	AAC76373	Human ORFX ORF1928
11	637.6	13.5	1624	23	AAF87118	NOV7 coding sequen
12	517.4	11.0	1577	23	AAF87121	NOV10 coding sequen
13	220.8	4.7	1522	21	AAAS7363	DNA encoding a hum
14	177.4	3.8	718	22	AAAS00854	Human cDNA clone H
15	109.4	2.3	722	24	AAAS62772	cDNA sequence #559
16	70.8	1.5	1414	13	AAQ27190	Tumour necrosis fa
17	70.8	1.5	1414	13	AAV71778	Tumour necrosis fa
18	70.8	1.5	1414	22	AAAD6019	Human tumour necro
19	70.2	1.5	1182	23	AAAS4364	DNA encoding novel
20	63	1.3	1411	22	AAH23114	Osteoarthritis tis
21	57.4	1.2	1354	22	AAAD08974	Alternative versio
22	56.6	1.2	985	21	AAAS7364	DNA encoding a hum
23	56.6	1.2	1365	22	AAI58121	Human polynucleoti
24	56.6	1.2	1984	20	AAAS2252	Protein PRO271 CDN
25	56.6	1.2	1985	22	AAAS5938	Human DNA encoding
26	56.6	1.2	1985	22	AAAF2410	Human PRO271 cDNA.
27	55.4	1.2	781	22	ABA08712	Human stabilin-I h
28	55.4	1.2	2808	21	AAAS9054	Human secreted pro
29	52.6	1.1	1851	22	ABA83182	HOSF-4 ovarian tum
30	51.2	1.1	11344	23	AAAS92464	DNA encoding novel
31	49.8	1.1	4480	19	AAV43590	Human epithelial g
32	49.8	1.1	4877	22	AAH28219	Nucleotide sequenc
33	49.8	1.1	5007	22	AAAS1657	Human polynucleoti
34	49.8	1.1	5032	22	ABA09190	Human precursor pr
35	49.8	1.1	5032	22	AAAS2641	Human polynucleoti
36	49.6	1.1	6567	11	AAQ03324	Elmeria tenella ge
37	45.4	1.0	1422	23	AAAS2463	DNA encoding novel
38	44.2	0.9	2895	22	AAI67200	Nucleotide sequenc
39	44.2	0.9	2897	22	AAH78221	Human bone marrow
40	42.8	0.9	436	22	AAK30592	Human secreted pro
41	42.2	0.9	1308	21	AAAS1263	Human cDNA SEQ ID
42	42.2	0.9	1640	22	ABA06583	Human Immunoglobul
43	42.2	0.9	1640	22	AAAS2787	Human cDNA SEQ ID
44	42.2	0.9	1641	22	ABA06754	Human Immunoglobul
45	42.2	0.9	1641	22	AAAS2848	Human Immunoglobul

## ALIGNMENTS

RESULT 1	ABAO4648	ABAO4648 standard; cDNA: 4706 BP.
ID	ABAO4648	
XX	ABAO4648:	
AC	22-FEB-2002 (first entry)	
DT		
XX		
DE	Rat Hyaluronic Acid Receptor for Endocytosis, HARE, coding sequence.	
XX		
KW	HARE, Hyaluronic Acid Receptor for Endocytosis; hyaluronan; chondroitin;	
KW	chondroitin sulphate; extracellular matrix; cartilage; skin;	
KW	vitreous humour; endocytic receptor; glycosaminoglycan; rat; ss.	
XX		
OS	Rattus norvegicus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..4296
FT		/*tag= a
FT		/partial
FT		/product= "Rat HARE"
FT		/note= "No start codon given"
XX		
PN	WO200181544-A2.	
XX		
PD	01-NOV-2001.	
XX		
PE	25-APR-2001; 2001WO-US13403.	
XX		
PR	25-APR-2000; 2000US-199538P.	
XX		
PR	02-NOV-2000; 2000US-245320P.	
XX		

PA (WEIG/) WEIGEL P A.  
 PA (ZHOU/) ZHOU B.  
 PA (WEIG/) WEIGEL J A.  
 XX  
 PI Weigel PA, Zhou B, Weigel JA;  
 DR WPI: 2002-049271/06.  
 DR P-PSDB: AAM47675.  
 XX  
 PT New mammalian hyaluronic acid receptor for endocytosis, useful e.g. for  
 PT identifying agents that inhibit binding to hyaluronic acid, and related  
 PT nucleic acid  
 PS  
 PS Claim 49: Fig 21; 263bp; English.  
 XX  
 CC The present invention relates to sequences for rat and human HARE  
 CC (Hyaluronic Acid Receptor for Endocytosis, ABA04648, ABA04662, AAM47675  
 CC and AAM47684). HARE can bind specifically to at least one of hyaluronic  
 CC acid (HA, also known as hyaluronan), chondroitin (C) or chondroitin  
 CC sulphate (CS). HA is an extracellular matrix component of all tissues,  
 CC in particular cartilage, skin and vitreous humour. HARE is the endocytic  
 CC receptor responsible for removing HA and other glycosaminoglycans from  
 CC the circulation. The present sequence is the coding sequence for rat  
 CC HARE.  
 CC  
 XX  
 XX Sequence 4706 BP; 1169 A; 1273 C; 1230 G; 1034 T; 0 other;  
 SQ

Query Match 100.0%; Score 4706; DB 24; Length 4706;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 4706; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 601 tggctgctctcttggcccaaatgtccaagctgccccgggagagtgcaaatgtgtgt 660  
 QY 661 ctgggaacggctctctgtctgggaacggtgtgtaatgcaactgtgacgtgcccagtgccggtg 720  
 Db 661 ctgggaacggctctctgtctgggaacggtgtgtaatgcaactgtgacgtgcccagtgccggtg 720  
 QY 721 ggtctcaatgggaacgctgtgaaacctgtgacatgagggggaagtatgttaccactgtgac 780  
 Db 721 ggtctcaatgggaacgctgtgaaacctgtgacatgagggggaagtatgttaccactgtgac 780  
 QY 781 caagcatgtctctgtgtccatgtggagatgtagccaagacccttgggagacggtcctgt 840  
 Db 781 caagcatgtctctgtgtccatgtggagatgtagccaagacccttgggagacggtcctgt 840  
 QY 841 gactgtgacgtcggctggcgagaggtgaaagtgtgacatgtgagatcaccacagaataatgtc 900  
 Db 841 gactgtgacgtcggctggcgagaggtgaaagtgtgacatgtgagatcaccacagaataatgtc 900  
 QY 901 aacggagcctgtgacacacacagtgccaactgtcctctgtgataccagacggaagcctctgtc 960  
 Db 901 aacggagcctgtgacacacacagtgccaactgtcctctgtgataccagacggaagcctctgtc 960  
 QY 961 aaatgtcgcgcaggaattccgagggaaatggaacggtctgcacagccatcatgtcgtgtgag 1020  
 Db 961 aaatgtcgcgcaggaattccgagggaaatggaacggtctgcacagccatcatgtcgtgtgag 1020  
 QY 1021 accagcaatggagagatgtttctacaagggcgcactgtaaagaaccaccccgagaacccg 1080  
 Db 1021 accagcaatggagagatgtttctacaagggcgcactgtaaagaaccaccccgagaacccg 1080  
 QY 1081 gtgtgtgtgtcgaagcgacgtctataccggcgacgcatcgtgtccttgaataataacccg 1140  
 Db 1081 gtgtgtgtgtcgaagcgacgtctataccggcgacgcatcgtgtccttgaataataacccg 1140  
 QY 1141 tgttggagaacacatgtgtgtgtgtgacagaaatgcaagatgcacacagagacaggcccaac 1200  
 Db 1141 tgttggagaacacatgtgtgtgtgtgacagaaatgcaagatgcacacagagacaggcccaac 1200  
 QY 1201 caggcgcgtctgtaactgtctgtcgaagtacacatgtgagatgtgaaaggtctgtcgtttac 1260  
 Db 1201 caggcgcgtctgtaactgtctgtcgaagtacacatgtgagatgtgaaaggtctgtcgtttac 1260  
 QY 1261 aatgtctgcttaacgaacaaatgtgcgcgtgcagttccattgctccttctgcaactacatgtag 1320  
 Db 1261 aatgtctgcttaacgaacaaatgtgcgcgtgcagttccattgctccttctgcaactacatgtag 1320  
 QY 1321 caagatccaaggatattgtaactgtgcaagccagactacacgggtgtgatggaatcgtctgcgg 1380  
 Db 1321 caagatccaaggatattgtaactgtgcaagccagactacacgggtgtgatggaatcgtctgcgg 1380  
 QY 1381 ggcagcatctacgggagacttcccaagaaccttgacgttcccaatcttcttcagttg 1440  
 Db 1381 ggcagcatctacgggagacttcccaagaaccttgacgttcccaatcttcttcagttg 1440  
 QY 1441 caggagcatgtcgtctcgagagactgtctggaacctgtgcccccttcaacggtgtctgcgcttg 1500  
 Db 1441 caggagcatgtcgtctcgagagactgtctggaacctgtgcccccttcaacggtgtctgcgcttg 1500  
 QY 1501 tctagctccttcaatcatgagcccggaattaaagactggagatcagcaaggcctcatgtcc 1560  
 Db 1501 tctagctccttcaatcatgagcccggaattaaagactggagatcagcaaggcctcatgtcc 1560  
 QY 1561 caggttcttcgtatacagtgatgtgggtgcgcagcagctgtctgttggacaacctaaagtg 1620  
 Db 1561 caggttcttcgtatacagtgatgtgggtgcgcagcagctgtctgttggacaacctaaagtg 1620  
 QY 1621 accacaagtgcaagacccctccaagagagcaggttccatctcttctcaagacat 1680  
 Db 1621 accacaagtgcaagacccctccaagagagcaggttccatctcttctcaagacat 1680  
 QY 1681 gtgtcataacatagcggaaggtccctgtcagtgatcatcatcagccaatgtgtgt 1740



Db 1681 gfttcataaacaatgagcgaaagttccctggtccagtacatcatcagcaaccaatgcgctc 1740  
QY 1741 atccacgtatatagacaagtctgtctcccaaaaacttgcttatcaaccccccaagaatgccc 1800  
Db 1741 atccacgtatatagacaagtctgtctcccaaaaacttgcttatcaaccccccaagaatgccc 1800  
QY 1801 ttggcagaagttctctgcaaaatcttactacagtggcagcaaacacagatataccaatctc 1860  
Db 1801 ttggcagaagttctctgcaaaatcttactacagtggcagcaaacacagatataccaatctc 1860  
QY 1861 agcaaatgttatacaaggaactcagaagcttgctgtcagttcaatcaactgaactccatccacca 1920  
Db 1861 agcaaatgttatacaaggaactcagaagcttgctgtcagttcaatcaactgaactccatccacca 1920  
QY 1921 gtcactgtctcttgctgcttaagcgaacaaagccctggaaagccttgcgcccaagacagcagac 1980  
Db 1921 gtcactgtctcttgctgcttaagcgaacaaagccctggaaagccttgcgcccaagacagcagac 1980  
QY 1981 ttccgttcaatcaagaacaagaacaagaagctgaaagtcttaccctgaaattccacgtgac 2040  
Db 1981 ttccgttcaatcaagaacaagaacaagaagctgaaagtcttaccctgaaattccacgtgac 2040  
QY 2041 cgaagactccaagagctttagcttcaagacctcccaagttctgcttccctgcaagaacctgcaaa 2100  
Db 2041 cgaagactccaagagctttagcttcaagacctcccaagttctgcttccctgcaagaacctgcaaa 2100  
QY 2101 ggtcacaagagcttgagttgtaggtgtggaactgacgtgacatcaggtgagttcttccaaac 2160  
Db 2101 ggtcacaagagcttgagttgtaggtgtggaactgacgtgacatcaggtgagttcttccaaac 2160  
QY 2161 gaacaacaatggtgcagatctacatacacgcgggagactctgtttgacgtggtgttgacctatgac 2220  
Db 2161 gaacaacaatggtgcagatctacatacacgcgggagactctgtttgacgtggtgttgacctatgac 2220  
QY 2221 attgactgctactcaatgaatccctacactaggttgacgtgacatcttactactctc 2280  
Db 2221 attgactgctactcaatgaatccctacactaggttgacgtgacatcttactactctc 2280  
QY 2281 gatattccgggggagtgacggaagtgcattcttaccctcccaaatgccccatggaaggagcaag 2340  
Db 2281 gatattccgggggagtgacggaagtgcattcttaccctcccaaatgccccatggaaggagcaag 2340  
QY 2341 ccaaaagggcgtgaaagaagaagtgtactctacaacccgttactactctcaaggaagaatgga 2400  
Db 2341 ccaaaagggcgtgaaagaagaagtgtactctacaacccgttactactctcaaggaagaatgga 2400  
QY 2401 ggcgtgcagaacactgtgcacggtgtgtatccaaccccccaagtgctgcacatggttactctc 2460  
Db 2401 ggcgtgcagaacactgtgcacggtgtgtatccaaccccccaagtgctgcacatggttactctc 2460  
QY 2461 atgccaagactgtccaagccctgacgcttgagaagcagaatacacggtgtataaacacccgggacatg 2520  
Db 2461 atgccaagactgtccaagccctgacgcttgagaagcagaatacacggtgtataaacacccgggacatg 2520  
QY 2521 ttgcgcgactctgttacaacaaccaatgggaacagtgctactatgcaacacggcttcaacgggaca 2580  
Db 2521 ttgcgcgactctgttacaacaaccaatgggaacagtgctactatgcaacacggcttcaacgggaca 2580  
QY 2581 gacctgcagagctctgcttgcacatgggaagtattggcctgactgtccagccccgcagctgtccc 2640  
Db 2581 gacctgcagagctctgcttgcacatgggaagtattggcctgactgtccagccccgcagctgtccc 2640  
QY 2641 gagcactgaacaagtgtgataagagggaatccaagagctccggaggatgtgctctgtataaacagg 2700  
Db 2641 gagcactgaacaagtgtgataagagggaatccaagagctccggaggatgtgctctgtataaacagg 2700  
QY 2701 ttggaacagccgactctgctgttacaactcccaagctgtatattcgcagtggtgcaacactgttcg 2760  
Db 2701 ttggaacagccgactctgctgttacaactcccaagctgtatattcgcagtggtgcaacactgttcg 2760  
QY 2761 ttccgtgcacgcacactgttacaaggagaacaacacggtgtgtgttgaacttgaactcagaaggt 2820  
Db 2761 ttccgtgcacgcacactgttacaaggagaacaacacggtgtgtgttgaacttgaactcagaaggt 2820

QY 2821 gacgggatalacatgacagctcgtgacatcttgcaaacaggaacaacagggggctgttgcgaag 2880  
Db 2821 gacgggatalacatgacagctcgtgacatcttgcaaacaggaacaacagggggctgttgcgaag 2880  
QY 2881 gtcgcataagtgctcccaagaaagcaccacaagtctcttgcagcttgcaagaagaagctacaag 2940  
Db 2881 gtcgcataagtgctcccaagaaagcaccacaagtctcttgcagcttgcaagaagaagctacaag 2940  
QY 2941 ggggagatgctacagctgcaatagaatagaacccctgtgcagacggtgttcaacggggagatgc 3000  
Db 2941 ggggagatgctacagctgcaatagaatagaacccctgtgcagacggtgttcaacggggagatgc 3000  
QY 3001 catggaacagccacccgcagagatgacgggccccaggccaagcataagtgtgaatgtaaaagt 3060  
Db 3001 catggaacagccacccgcagagatgacgggccccaggccaagcataagtgtgaatgtaaaagt 3060  
QY 3061 caatagttcgggagcgaatgagctgtgagcttgagcaactgcgcgtcgaacgctgtacta 3120  
Db 3061 caatagttcgggagcgaatgagctgtgagcttgagcaactgcgcgtcgaacgctgtacta 3120  
QY 3121 caggaacaacggaacagtgccaccagaatgcccagctgtgcagacctacttccaagaacag 3180  
Db 3121 caggaacaacggaacagtgccaccagaatgcccagctgtgcagacctacttccaagaacag 3180  
QY 3181 accgttaggaggtattccatctacgctcccaactgggcccagttacaactatgtgaca 3240  
Db 3181 accgttaggaggtattccatctacgctcccaactgggcccagttacaactatgtgaca 3240  
QY 3241 gccaaagaagcctgtgccaagaagctgacacatagccacactacaacacagctctctat 3300  
Db 3241 gccaaagaagcctgtgccaagaagctgacacatagccacactacaacacagctctctat 3300  
QY 3301 gcccaagaagcccaagtatacactgtgtcgcgcgcgtgacgtgagaggtgagcggtgtgcc 3360  
Db 3301 gcccaagaagcccaagtatacactgtgtcgcgcgcgtgacgtgagaggtgagcggtgtgcc 3360  
QY 3361 taacccgaactagttatgacctctcagaaggttggtgccaagctgttttggaatctgtaagatc 3420  
Db 3361 taacccgaactagttatgacctctcagaaggttggtgccaagctgttttggaatctgtaagatc 3420  
QY 3421 ggaatccagggccacaagaagatgaatgttggaatgtcttctgttaccggaatgaagaatgtg 3480  
Db 3421 ggaatccagggccacaagaagatgaatgttggaatgtcttctgttaccggaatgaagaatgtg 3480  
QY 3481 aactgcaacctgcaaggcaggtcatgttggaatgtgcttctcgttgcaagtggaaactgtctg 3540  
Db 3481 aactgcaacctgcaaggcaggtcatgttggaatgtgcttctcgttgcaagtggaaactgtctg 3540  
QY 3541 caggtctctcatgttcccttcccctcgcctcacaacattcccttgacaagaaggtgtgtgttcttc 3600  
Db 3541 caggtctctcatgttcccttcccctcgcctcacaacattcccttgacaagaaggtgtgtgttcttc 3600  
QY 3601 aagaagctcagcccgagggaagaagcatcttctgaacaacacttaccagactgtgccatccgtgac 3660  
Db 3601 aagaagctcagcccgagggaagaagcatcttctgaacaacacttaccagactgtgccatccgtgac 3660  
QY 3661 accctgttctgtgccaacagaacaagtgggtctaacccgggaataaagaacctgtctgcccggagc 3720  
Db 3661 accctgttctgtgccaacagaacaagtgggtctaacccgggaataaagaacctgtctgcccggagc 3720  
QY 3721 atttggaacacacactcaatgttcaacgtctccttcttaaatgaaactgttcaatgttgaatc 3780  
Db 3721 atttggaacacacactcaatgttcaacgtctccttcttaaatgaaactgttcaatgttgaatc 3780  
QY 3781 ttcttgaagactatgtctgggaagccaactgtcattacccttaccgcaaggaacagctccac 3840  
Db 3781 ttcttgaagactatgtctgggaagccaactgtcattacccttaccgcaaggaacagctccac 3840  
QY 3841 caaaggaacaggttctgtggaatgaagaatccattcttgcaggtggaacataatgcgcgaat 3900  
Db 3841 caaaggaacaggttctgtggaatgaagaatccattcttgcaggtggaacataatgcgcgaat 3900

QY	3901	ggatctctccataattaattctbaaccttggaaagctctcccaaggagcaaggctggcc	3960
Db	3901	ggatctctccataattaattctbaaccttggaaagctctcccaaggagcaaggctggcc	3960
QY	3961	cactctcgccctggggagcaagtaatactctgtgcgcgtctctgttaacctggtgcattgct	4020
Db	3961	cactctcgccctggggagcaagtaatactctgtgcgcgtctctgttaacctggtgcattgct	4020
QY	4021	ctggcagcttaactcttactcttcggtctaagaagcgaaaccaactggtttccacggtttgat	4080
Db	4021	ctggcagcttaactcttactcttcggtctaagaagcgaaaccaactggtttccacggtttgat	4080
QY	4081	cagaagaagacatttgaatgtctcttgaccttttggcaagaagacgcccagaagatctgcaaac	4140
Db	4081	cagaagaagacatttgaatgtctcttgaccttttggcaagaagacgcccagaagatctgcaaac	4140
QY	4141	ctctgtatgagaccctcagcgcgccagccccaagtcctctctgtgacccttcacagacc	4200
Db	4141	ctctgtatgagaccctcagcgcgccagccccaagtcctctctgtgacccttcacagacc	4200
QY	4201	ctggagaagaagatctcgagagaagcgaaacctctgggggcaattggtgctctggaactgga	4260
Db	4201	ctggagaagaagatctcgagagaagcgaaacctctgggggcaattggtgctctggaactgga	4260
QY	4261	agccagaagaacaacaacagatcgaatgcttccacggtgaattccagcccacagctgtccatg	4320
Db	4261	agccagaagaacaacaacagatcgaatgcttccacggtgaattccagcccacagctgtccatg	4320
QY	4321	atcagttgttttaagaatgatacaacatcataaagccagccatactcaacctctgtgta	4380
Db	4321	atcagttgttttaagaatgatacaacatcataaagccagccatactcaacctctgtgta	4380
QY	4381	atctcggattgtgcgaaggcttaagaagcaatgtgctcgtgatacttgaggggaaacctcac	4440
Db	4381	atctcggattgtgcgaaggcttaagaagcaatgtgctcgtgatacttgaggggaaacctcac	4440
QY	4441	ctctcccgagccctaaacagtggtctctcctaactccatatgagtgctgtgctgtcttgcc	4500
Db	4441	ctctcccgagccctaaacagtggtctctcctaactccatatgagtgctgtgctgtcttgcc	4500
QY	4501	tctctgttaaccacaacactgtgaactctgtgtgatacttccctatitgaagtaagcaacaagg	4560
Db	4501	tctctgttaaccacaacactgtgaactctgtgtgatacttccctatitgaagtaagcaacaagg	4560
QY	4561	cggggcttaaccttctatgtgtctgtatctcaagtaagcaagaatgaactctgccaacatggt	4620
Db	4561	cggggcttaaccttctatgtgtctgtatctcaagtaagcaagaatgaactctgccaacatggt	4620
QY	4621	gtctaataaattgttttggagaacaataaagaagcactgtgtlaacctagaagtgtaaa	4680
Db	4621	gtctaataaattgttttggagaacaataaagaagcactgtgtlaacctagaagtgtaaa	4680
QY	4681	ctatgaaagcaaaaaaaaaaaaaaa 4706	
Db	4681	ctatgaaagcaaaaaaaaaaaaaaa 4706	

RESULT	2	
ABAB04662		
ID	ABA04662	standard; cDNA; 4576 BP.
XX		
AC	ABA04662;	
XX		
DT	22-FEB-2002	(first entry)
XX		
DE	Human Hyaluronic Acid Receptor for Endocytosis, HARE, coding sequence.	
XX		
KW	HARE; Hyaluronic Acid Receptor for Endocytosis; hyaluronan; chondroitin;	
KW	chondroitin sulphate; extracellular matrix; cartilage; skin;	
KW	vitreous humour; endocytic receptor; glycosaminoglycan; human; ss.	
XX		
OS	Homo sapiens.	
XX		

Key	Location/Qualifiers
CDS	1..4185
FT	/**tag= a
FT	/partial-
FT	/product= "Human HARE"
FT	/note= "No start codon given"
XX	
PN	MO200101544-A2.
XX	
PD	01-NOV-2001.
XX	
PF	25-APR-2001; 2001WO-US13403.
XX	
PR	25-APR-2000; 2000US-199538P.
XX	
PR	02-NOV-2000; 2000US-245320P.
XX	
PA	(WEIG/) WEIGEL P A.
XX	
PA	(ZHOU/) ZHOU B.
XX	
PA	(WEIG/) WEIGEL J A.
PI	Weigel PA, Zhou B, Weigel JA:
XX	
DR	WPI; 2002-049271/06.
XX	
DR	P-PSDB; AAM47684.
XX	
PT	New mammalian hyaluronic acid receptor for endocytosis, useful e.g. for
XX	identifying agents that inhibit binding to hyaluronic acid, and related
PT	nucleic acid -
XX	
PS	Claim 49; Fig 33; 263pp; English.
XX	
CC	The present invention relates to sequences for rat and human HARE
CC	(Hyaluronic Acid Receptor for Endocytosis, ABA04648, ABA04662, AAM47675
CC	and AAM47684). HARE can bind specifically to at least one of hyaluronidic
CC	acid (HA, also known as hyaluronan), chondroitin (C) or chondroitin
CC	sulphate (CS). HA is an extracellular matrix component of all tissues,
CC	in particular cartilage, skin and vitreous humour. HARE is the endocytic
CC	receptor responsible for removing HA and other glycosaminoglycans from
CC	the circulation. The present sequence is the coding sequence for human
CC	HARE.
CC	
XX	
SQ	Sequence 4576 BP; 1127 A; 1193 C; 1203 G; 1053 T; 0 other;
Query Match	60.2%; Score 2835.2; DB 24; Length 4576;
Best Local Similarity	79.2%; Pred. No. 0;
Matches 3483; Conservative	0; Mismatches 898; Indels 19; Gaps
Oy	67 attcatatacaactgycgaagtgcgaatcgagctgtgcagatgattactgttttcgacga 126
Db	
	1 attcataataatcttgccgatgcaatggatctggcgtgcgaatgccacaacagtglttgcacca 60
Oy	127 aacaatgaagaaccatcgaaaactatatcaggaggaagaagccacatctctaagaagaaatg 186
Db	
	61 aacaacaatgcccattcgagaatatacatcacaggaggaagaagatctgtctcttaaggagagac 120
Oy	187 attctacggtlaccatgttggtctctgggggaaaaagctctcctaagaagaatgcatgaaggac 246
Db	
	121 gtctctccgtatcatgtgtgtctcgtggaggaagaacctctgaagaatgacctgcacaatgac 180
Oy	247 atgcacagagaagacatgtctggggtctctcctactcctgtgcctcttcttcgcgaatgac 306
Db	
	181 atgcactggaagccatgtcgtgggttctctctacttctcttaagctctcttcctccaatgac 240
Oy	307 cagcttatgtataaigaagctccaataactataccaacatgtggcactgataaagaagtg 366
Db	
	241 cagcttcaatgaatgaagagctccaataactataccaacatgtlagcactgataaaggagtg 300
Oy	367 atccatgtgtctggaagaagattctcgaaatlcagaagaacagatgtgacataaataagcac 426
Db	
	301 atccatgtgtctggaagaagattctcgaaatlcagaagaacagatgtgatabataaataagcac 360
Oy	427 attattgtgagagagtgatgttgtaaagtgttccacgaagaagccccctgcacatcggagaca 486

[illegible]

Db	1440	tgacagccttgatgaggaagctcgggtlaaagacttvgagcaataacggtttaatgccc	1499
Qy	1563	ggtcttcgctacacgtatgtggcctgcgcagcagctgcgtttggacaacctaaagtgc	1622
Db	1500	ggttcttcggtaaccatgtgtgtcgcctgtccacagcttgcttcttggaaaacctgaattgt	1559
Qy	1623	cacaagtgccacgacccctcccaaggagagcagtttccatctctgtcttcagagacatgt	1682
Db	1560	ctcaaatgctacttcctccctcaaggagagcaatagtatctctcgtcttcagagcaggt	1619
Qy	1683	gtctataaacatctvgagcggaaggtccctgtccagtgatcatcatcagcaaccatrgcgtcat	1742
Db	1620	gcatataaacaaataagagctaaagatcatatccagtgatcatcatcaagtaactatcaggtatgt	1679
Qy	1743	ccagcttaatagaacaagtgtctgtccccaacaacttgctatcatcccccacaagaatgcct	1802
Db	1680	tcatatcatagacaatatgtctatctcccaaaaatttgctatcatctcccaagacaactc	1739
Qy	1803	gggcaggggttctgcacaacatctactacagtgtagcaagcaaacccagatatacagaaattcag	1862
Db	1740	tggagaattctgcacaacatcttaacgactttggcaacaacaatgtcatcatcaaatlttag	1799
Qy	1863	caagttgatacagcgactcaagctctgtctgtcaatcatcatctgtcatccatccaccccaat	1922
Db	1800	caacttaatacagagctcagagtttctgtcgtatgtcatcaccgatcccatccaccccaagt	1859
Qy	1923	cacttcttcttgccctacagcaacaagcccttgaaagccttgcccacagagcagcagact	1982
Db	1860	cactctctcttgcccaaccgaccaaagccctccatgcctactactgtctgtaacaacaggaact	1919
Qy	1983	cctgttccaatccaagacaacaagagacaaagcttgaaagtcttcaactgtaagttccacgtgtatccg	2042
Db	1920	cctgttccaacccaagacaacaagagacaaagcttgaaagagtatttgaaagttcatgtatagctg	1979
Qy	2043	agaccccaagcgtttagctgtccagactcccccagctcgtctctccctvgaaagaccctgcaag	2102
Db	1980	agatgccaaaggttttagcctgtgtgactcttcccacatccactgtccttggaaagacccttgaag	2039
Qy	2103	ctccagagctgaattgtgaaagtgctggaacatcgcagtgacatccgcttgagctctttctaaaga	2162
Db	2040	ttcagagctcgaatgtgaaatgtggaagctcgcgcagggacatccggtugaccccttctctgaaatg	2099
Qy	2163	acaaaatgtgcagatcatatacaaccgcggagactctctgttgaacggtgtgtgtgacctatgacat	2222
Db	2100	ccaaaacctgcagaattgtgtgcagcgggagctctgtttagccctgggtgtgtgacctatgacat	2159
Qy	2223	tgactgcctactcatgatcatctcaaccctaaagtgagccgaatgtagaacctttactactctga	2282
Db	2160	tgactgtctgcgtgatgtgactcccaacctcgtgggcgcgtctgtgacactcttactactctga	2219
Qy	2283	tattccggggagagtcgcgaaagtgtgacttttcaactccaaatgtgcccactgaaagaagacgc	2342
Db	2220	tgccctcgggggagtgatggaagctgtgtgtcaataatcccaagctgcgcacaagtgtgaaataaac	2279
Qy	2343	aaaagtgcttgaagaagaagtgtatctatacaaacccgttactcttcagagaggaacgttggaaag	2402
Db	2280	aaaaggtgtgtgaagcagaagtgtctctataacc--tgccctcaagaaggaacacttggaaag	2336
Qy	2403	ctgcacagaacctgtgtcacccgtgtgtatctcaaaaccccccaggtgtgcgtccatgttactcat	2462
Db	2337	ctgtccgggagcgtgttcaacacctgtgtatctacaagatctcccaggtgtctgtcaaggtgcactatccg	2396
Qy	2463	gccaacactgtcaagcgtcgtccctcggggagcccaatacacccgtgttaacaacccgggcatgtg	2522
Db	2397	gcgaacactgtcaagcgtcgtccctcggggagcccaatgccccgtgttaataaacccgggtgtgtctg	2456
Qy	2523	ccgcagactgttacaacaccatgtgaaagctgtcccttagtgcacaacccggtcttcaacgggagacac	2582
Db	2457	ccgttagctagtaactgtgcacaccggagagtgtgaatgtgcaaacaccggtcttcaatvggaaagcgc	2516
Qy	2583	ctgcagactctgtgcagcatgtggagaatttgggcctgtacgttcaagcccgcagctgtccctga	2642
Db	2517	gtgttaagatgttcctgcgcgggagaagatttggagccttatgtctcgcctcgtgtgagctgtccaca	2576

QY	2643	gcaevgaacagvtgtatgtagagvgatatacaaggtccvgvgagvtgcctctgtgaaacagvggt	2702
Db	25777	ccaacgagacatgtgcgaatgtatgtcatatacvggtcccgvgcagvtgcctctgtgaaacagvggt	2636
QY	2703	gacagccgtctgtgtgaacatcccaacagcttatttcgacgtgtgcacaccgtctgtcc	2762
Db	2637	gacagggccctctgtgtgatacctccaagcagtttctgtcgtgagvtgtacgcctctgtcc	2696
QY	2763	ctgtacacgccaactgttaccgagagaacaacacgtgtgtgtttaacttgcacccaagvtga	2822
Db	2697	tgctcatgtccacctgttaagvagagaacaacacgtgtgtgtttaacttgcataatgtgaagtga	2756
QY	2823	cvgggtaccatgtgcacagttcgtgtgacttctgtgaaacagaaacaaacggvggtgtgtgcgaagt	2882
Db	2757	cvggaatcaatgtgcacagttcgtgtgacttctgtgaaacagaaacaaacggvggtgtgtgcgaagt	2816
QY	2883	ccgttaagtgtctcccaagaagvgacccaagctctcttgcacgtctgcagaagaaagvtacacaag	2942
Db	2817	ggccagaagtgtctcccaagaagvgacccaaggtctctctcgtcagctgcagaagaaagvtatacaag	2876
QY	2943	ggatgtggtctacagctgtcatatagagataagacccctgtgtcagaacggtgttacaacggvggtaccca	3002
Db	2877	ggacagvgcacagctgtcacacgagataagacccctgtgtcagaacggtgttacaacggvggtaccca	2936
QY	3003	tgagacacgccaactgtcagagataagacggcccaagccaagtcataagtgtgaatgtlaaagaagtca	3062
Db	2937	cqagacacgccaactgttgaatagaatgacagggcccgvgcaagacaaagtgtgagvtgtataaagtca	2996
QY	3063	ctatgtctcgvggagacggatgtgactgtgagcctctgagcaagctgtccgctcgacccgttgttaca	3122
Db	2997	ctatgtctcgagagatgtggtctgaactgtgtagacggvgacagctgtcccatgtgacacggctgttaca	3056
QY	3123	ggacaaacggacaggtgtgcacccaagatgtccagcgtctgtgagacgtcttacttccaagaacagac	3182
Db	3057	ggacaaatvggcaggtgtgcacacgacacgacaaatgtgtctcgacatctccacttccagagataccac	3116
QY	3183	cgtaagagagatttccatctacgtctcccaacttgggcagtcataacatgtacatttgcacaagc	3242
Db	3117	tggtgggtgttccatctacgtctcccaacttgggcagtcataacatgtacatttgcacaagc	3176
QY	3243	caaagagagcgtgtgccaagaagaactgtgcggccatatagcacacttaacaacagctctccatgtc	3302
Db	3177	cagagagagcctgtgcacaacgaagaactgtgcagacaaatgcacacttaacaacagctctccatgtc	3236
QY	3303	ccaagaagccaagtatacactctgtgcctcggccggcctgtgcctgagagatgtgvcggttgtccata	3362
Db	3237	ccaagaagccaagtatacactctgtgcctcggccggcctgtgcctgagagacggvgggttgtccata	3296
QY	3363	cccggataatgtatgctcttcagagaagtgtgtgtgaaacgttgtgtggatctgttaacatacgg	3422
Db	3297	cccgaagaagccttgcgtctccagaacacgtgtgcctctgtgtgtgtgtgtgtatgtatgtacatacgt	3356
QY	3423	atccagagcgcaacaagaagtgaatatgtgtgagatgtcttctctgttaccggatagaaagaatgtgaa	3482
Db	3357	actctagaaccaagaagaagtgaatatgtgtgagatgtcttctctgttaccggatagaaagaatgtgaa	3416
QY	3483	ctgacacctgcagaagcgaggtatgtgtvggagatgtgctctcgtgtcagatgtvggaacctgtgtca	3542
Db	3417	ctgacacctgcagaaggtgtggtatgtgtvggagatgtgcttctctcagatgtvggaaacctgtgtca	3476
QY	3543	gggtcccatatgtcttccctccgttcaaaaacttccctgcagaagaaggtgtgtgcttctttccaa	3602
Db	3477	gggtccctgtatgtcttccctccactcaacaacttccctgcagaagaagvtgtgtgcttctttccaa	3536
QY	3603	gaactcagcccggaggaacaggtcttlttggaaacactgtacgtgacatgtccatccgtgtgcac	3662
Db	3537	cagctcagctccggvgccggtcatlttctagaagaacctgtacgtgacatgtccatccgvgcac	3596
QY	3663	ccgtgttgtgtccacaagaacatgtgtggtctacccgggaataagaagcgtgtctgtccggvgacat	3722
Db	3597	ccctcttgtgtccacaagaacatgtgtggt	3656

Chr	Start (kb)	End (kb)	Gene	Transcript	Sequence
QY	3723	3782	tgagcaccactcaatcatgtcaacgctcccttctaagaatgacctgttaatgtaacct	3782	
Db	3657	3716	cgagaccacacttcgcacatgtcaacgacgttttcttacaatgacctgtcaatggaaccac	3716	
QY	3783	3841	tctgagacctatgcctggagaagccaaactgcaccttaacctttagcgaagacaaagctccaac	3841	
Db	3717	3776	cctgcaaacagagcctgggaaagcaagcgtctatcaactcgcagccagacgaacctccaacc	3776	
QY	3842	3899	--aagaagccaggtttggaatggaagatccatctcgaatggaatcatcatcgcgcacaa	3899	
Db	3777	3836	gacggagagccaggttctgttgaatggaagagccattctcgaatggagacattcttcccca	3836	
QY	3900	3959	tggaatcccatatatttctctgaacctttgagaagctccctcccaacggagcaacggctgc	3959	
Db	3837	3896	tggagatcatcatgtcatcttccaaagcctttaaagacaccccttgcgcccttgac	3896	
QY	3960	4019	ccaaacttgccctggggaagatattatcttgcgcgtgtccctgcctgaactgtgtgcatltag	4019	
Db	3897	3956	ccaaacttgccctggggaagatattatcttgcgcgtgtccctgcctgaactgtgtgcatltag	3956	
QY	4020	4079	tcctgacacttaactcttcaactcgcgtcaagaacgacgacacactggtttccagcgtttga	4079	
Db	3957	4016	cttgctgtcttaactcttcaactcttgcataaaccggaacaatcgctctcagcatltag	4016	
QY	4080	4138	-tcagaagaagacatltgatgtcttgaccttttgcaagcagcagcccaagaatatatgcgaa	4138	
Db	4017	4076	gtcggagaagagacatlaatgtltgcagctctcttgcaagcagcagccttgaatatctcgaa	4076	
QY	4139	4198	ccctctgtatagacctatagcgcgcgcaccccaagagctctctcctgttacccttcaacga	4198	
Db	4077	4136	ccctctgtatagacctatagcgcgcgcaccccaagagctctctcctgttacccttcaacga	4136	
QY	4199	4255	cccttgagaacacggatcttgagaagacgacacctcttgagggacatgc--gctctgaca	4255	
Db	4137	4196	ctcttgagaacacgacgcttgagggaatgacacctcttgaggacactgtgaggccttgagc	4196	
QY	4256	4315	tgagaagccagcaagaacaccacagtcacggttccacaggtgatctccagccccagctgtct	4315	
Db	4197	4253	ggaagatgcacatcaactcaactgcacactcggccatcaactgtgaattctcagcac--	4253	
QY	4316	4375	catgtgacgtgtgtttaagaatgacaacactcaataagcagccatcaactcaacctct	4375	
Db	4254	4313	catgtgacgtgtgtttaagaatgacaacactcaataagcagccatcaactcaacctctctct	4313	
QY	4376	4435	ggttaactcctggatctgtcgcgcagggctaaagagccatgttgcctgtgaatacctggggacc	4435	
Db	4314	4372	ggttaactcctggatctgtcgttctctgtggagagatgtgtg-ctgtgccaccagataca	4372	
QY	4436	4455	tcacactcctctgaagcctat	4455	
Db	4373	4392	gctctcctcctgcacacctt	4392	
RESULT 3					
AAf87120					
ID	AAf87120 standard; DNA: 3625 BP.				
XX	AAf87120;				
XX	26-MAR-2002 (first entry)				
XX	NOV9 coding sequence.				
XX	NOV: Cytosolic; contrapeptive; antiinflammatory; immunomodulatory;				
KW	cardiovascular; casein kinase II phosphorylation site; contraception;				
KW	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;				
KW	epidermal growth factor; cell development; apoptosis; cell adhesion;				
KW	growth migration; cell structure; motility; cancer; immune disorder;				
KW	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;				
KW	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;				
KW	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;				
KW	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.				
XX					



```

||||| |||| | |||| | ||||| ||||| ||||| ||||| |||||
Db 1704 gacagtcgatagtatgcatcacagcgctccggcgagctgcctctgtgaaacggggtgagcaag 1763
Oy 2708 ccgcttgctgtgacacccccaacagctgtattcagaggtgtgacacccgctgtgcctgtgc 2767
Db 1764 gccctctgtgtgacacaaagcagtlctgcctgtcagatgtgtlaacgtccctcctgtctgtc 1823
Oy 2768 acgcacacgttaccggaagaaacacacgtgtgtgtlaactgtgaactgaagaaagtgacggaa 2827
Db 1824 atgcacacgttgaagaaacacacacgtgtgtgtlaacgtgtattatgaaggtgacggaa 1883
Oy 2828 tcacatcacagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2887
Db 1884 tcacatcacagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1943
Oy 2888 agtgcctcccgaaagagcaccacaaagtctctgtgcagctgcaaaaagaggttacaagggagtg 2947
Db 1944 gatgtcccgaaagagcagaaaggtctccgtgcagctgtccaaagaggtatcaaaagggagcg 2003
Oy 2948 gctacacgtgataagataagaccctgtgtgacagcggtgtacaagggaggtatgccatgagc 3007
Db 2004 ggcacagctgtgacacagataagaccctgtgtgacagcgcttacaagggaggtgtacagagc 2063
Oy 3008 acgcacacgtgacagatgacagggcccgagcaagcaatgtgtgtgtgtgtgtgtgtgtgtgtgt 3067
Db 2064 acgcacacgtgtgaagatgacagggcccgagcaagcaatgtgtgtgtgtgtgtgtgtgtgtgtgt 2123
Oy 3068 tcggagcgaggtgtgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3127
Db 2124 tcggagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2183
Oy 3128 acgagacagtgccacacagatgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3187
Db 2184 atgagcagctgtgcacacagcagcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2243
Oy 3188 gagatctccatctcgtctccctcactgtggccagtacaactgtgtgtgtgtgtgtgtgtgtgtgt 3247
Db 2244 ggggtgtccatctcagcctccctcactgtggccagtacaactgtgtgtgtgtgtgtgtgtgtgtgt 2303
Oy 3248 aagcctgtgccaagaagagctgtgcacacacacacacacacacacacacacacacacacacacac 3307
Db 2304 aggcctgtgccaagaagagctgtgcacacacacacacacacacacacacacacacacacacacac 2363
Oy 3308 aggcacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 3367
Db 2364 aggcacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 2423
Oy 3368 ctacgtatgctctcagaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3427
Db 2424 cagcctctgcctcccaagaacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2483
Oy 3428 gggcacaacaagatgaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3487
Db 2484 gaccacaacaagatgaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2531
Oy 3488 cctgtcaagcaggtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3547
Db 2532 ----- 2531
Oy 3548 tcagtctctccctcgtctacaacacttcctgacagaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 3607
Db 2552 -----gaaggtctgtgcctatccacaagct 2556
Oy 3608 cagcccgagacagcatttltgaaacacacacacacacacacacacacacacacacacacacacac 3667
Db 2557 cagctcgagcgcggtcatttctagaacacacacacacacacacacacacacacacacacacacacac 2616
Oy 3668 ttgtgtcacaagaacagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3727
Db 2617 ttgtgtcacaagaacagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2676
Oy 3728 accacctacacaaatgtcaaggtctccttttaacagacctgtcaatgtgtccttctgca 3787
||||| | ||||| | | | ||||| ||||| ||||| |||||

```

```

Db 2677 accacctcgccaatgtlcatgatttcttcaaatgaactgtgtcaatgtgcacacacctgtc 2736
Oy 3788 gacatagcttgggaagccaactgtctcatlaacttcagccaggaacacacacacacacacacacac 3844
Db 2737 aacagagcgctgggaagcaagctgtgtcatlaacttcagccaggaacacacacacacacacacacac 2796
Oy 3845 agacaggttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3904
Db 2797 agacaggttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2856
Oy 3905 tctccatattatttctgaacacttggagactcctctccacggagcaagcagctgtgcacac 3964
Db 2857 tcatcattcatattccagagcctttaaagcaacccctgtccctgtgaactgtgaactgtgaactgtga 2916
Oy 3965 ctgagcctggagacagatatactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4024
Db 2917 ctgagcctggagacagagatctcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2976
Oy 4025 cagcttactcttactctcgcgttaagcagcagacacacacacacacacacacacacacacacacac 4083
Db 2977 ctgcttactcttactctcgcgttaagcagcagacacacacacacacacacacacacacacacacac 3036
Oy 4084 aagagagacatgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4143
Db 3037 aagagagacatgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3096
Oy 4144 tgtatgagacactcagcgccgcaaccccaagagctcctctgtgtgtgtgtgtgtgtgtgtgtgtgt 4203
Db 3097 tgtatgagagacacactcagcgccgcaaccccaagagctcctctgtgtgtgtgtgtgtgtgtgtgtgt 3156
Oy 4204 gagaacacagatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4260
Db 3157 aagaacacagatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3216
Oy 4261 agcagcaagcaacacagatcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4320
Db 3217 tgcacagcatatcactcactgcacacacacacacacacacacacacacacacacacacacacac 3273
Oy 4321 atcagctgtttaaagaatgacaacactcataagcagacacacacacacacacacacacacacac 4380
Db 3274 gactttagaagcaatgaagctctttaaagcactacagaagcactacacacacacacacacacac 3333
Oy 4381 atctggagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4440
Db 3334 atctggagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3392
Oy 4441 ctctctgaagcctat 4455
Db 3393 ctctctgaagcctat 3407

RESULT 4
AAF87119
ID AAF87119 standard; DNA: 2483 BP.
XX
AC AAF87119;
XX
XX 26-MAR-2002 (first entry)
XX
DE NOV8 coding sequence.
XX
KW NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
KW cardiovacular; casein kinase II phosphorylation site; contraction;
KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
KW epidermal growth factor; cell development; apoptosis; cell adhesion;
KW growth migration; cell structure; motility; cancer; immune disorder;
KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.
XX
OS Unidentified.
XX

```

[illegible]

Db	631	ctcgtaaccabtgylgctgcctgcaccacgctcgtctcttggaaaactgaaattgatctca	690
Oy	1628	gtgcacagacccctccaaaggagacagtttccatctctctctccaggaactgtgtca	1687
Db	601	atgctactctccctcccaaggagagccaatagtcattccgcgtctctccaggaacggtgtata	750
Oy	1688	taaacabtgagcggaaggtccctgtcccaatgacataccagaccaatggcgtcatccacg	1747
Db	751	taaacataaagctaagaatcatatcatccagtatcatcatcagtaactaatgtgattgtcata	810
Oy	1748	ttatagacaagttgctgtctctcccaaaaactgtcttatcaaccctccaaagatgctctggga	1807
Db	811	lcatagacaacaattgctctctcccaaaaattgtcttatcatctcccaaaagacaacttggaa	870
Oy	1808	gggtcttcgcaaatcttactacaggtggcgcacacacagagatatactacaattccagcaat	1867
Db	871	gaattctgcacaactcttaagcaactttggcaacaaacaaatggctatcaacaatttaagcaact	930
Oy	1868	tgatacagagactccagctctgtctgtccagtcacatctactccatcccaaccacagctacgt	1927
Db	931	taatacagagactcaaggttctgtcgtggtatcaaccgatacccatccacaccacagctactcc	990
Oy	1928	lctctctgctactagacacaagcccctggagaactctgtcccccagagcagcaggaactctctgt	1987
Db	991	lctctctgccccacgcagacaagcccctccatgcctactacgtctgagacaacaagactctctgt	1050
Oy	1988	lcaatcaagagacaanaagacaagaactgtaagcttactctgaagtcttcaacgtgaltccgaagt	2047
Db	1051	lcaaccagagacaacaagagacaaagcttgaaagatctattctgaagtcttcaatgatacgaatgt	1110
Oy	2048	ccaaagctcttgacgtccacacccctccacgggtctcgtctctccggaagacactgcacaagctcag	2107
Db	1111	ccaaaggtttttagctgtgtagatcttcccaacatccacatgcgtcgtggaagacactgcacaaggttcaag	1170
Oy	2108	agctcgagctgtgaaggtgtgtggaactgtgcagatgacataccgtgtagactcttcttcaaacgaca	2167
Db	1171	agctcgagctgtgaaggtgtgtggaactgtgcagatgacataccgtgtagactcttcttcaaacgaca	1230
Oy	2168	tgctgcagattctacataccacggggaactctgtgtttagcgttggtgtgtgctctatgctatgact	2227
Db	1231	ccgtgcagaaattgtgtcagcgggaagactctctgttgaactcgggtgtgtggtcttcaagcaatgact	1290
Oy	2228	gctcaactcatgatactccataccactagatgtgcagatgtgtacacttctacactctcgatctac	2287
Db	1291	gtctgcgtcagatgtatccccaaccctgggggcgcgcgtgtgaacacttctactactctcgatgctc	1350
Oy	2288	cgaggggagctgcggaagttgcatcttccaactcccaatgcccactgaaagagagccacaag	2347
Db	1351	cgaggggagctgtcggaagctgtgtccaatactccacgtcccaaggtgtgagttaaacacaag	1410
Oy	2348	gcgtggaagaagaatgtatctatacaaccggttactcttcaaggaggaacgtgtggaagctgtgc	2407
Db	1411	gtgtggaagcagaatgtgtctctacaacc--tgcaccttcaaggaggaacactgcggaagctgtgc	1467
Oy	2408	agaaacctgtgtacacgggtgtgtatccaaacccccacgggtgtgtcgtccatgtctctcaatgcag	2467
Db	1468	ggggagcgtatgtcagactgtgtatacagatacccccaggtgtgtcgtcgaagggctactcttcgggcag	1527
Oy	2468	actgtcagagctctgcctcgtgagagacagataccacgtgtatacaaacccggggatgtgtgcgcg	2527
Db	1528	actgtcagagctctgcctcgtgagagacagatgtcccgtgtataacccgggtgtgtctctctgt	1587
Oy	2528	atctgtcaacaaccccatgtggagacagttgtctatgacacacccgggtcttcaacgggaagacactgcg	2587
Db	1588	atctgtcaactccgcacacccgggaagatgtataatgtcaacaacccgggtcttcaatgtggaacggtgtgt	1647
Oy	2588	agctcctatgtcgaatgcggaagatcttgggcctgtacactgtacagaccccgacgtctctccagcaat	2647
Db	1648	agatgtcgtcgtccggggagaagatttgggcctgtgattgtctgtccctgtgtcgtctcaagcaacg	1707
Oy	2648	gacacgtgtgataagggatctcacagactccggggagtgtcctgtctgtgaacaagggtgtgacag	2707
Db	1708	gacacgtgtgataagggatctcacagactcccggaaggtgtcctgtgtgaacaagggtgtgacag	1767



QY	2708	ccggttcgtgtgtaaacatcccaacagctglatctgcagagtgtaacactgtctgtccgtgc	2767
Db	1768	gccctctgtgtgaacaactcaagcagcttctgtctgcagtgtaacgcctccctgtctccgc	1827
QY	2768	acgcacactgtgacggaacaacaacgcgtgtgtgtgtaactgtgaactacgaagtgacggga	2827
Db	1828	atgcacactgttaaggagaacaacaacgcgtgtgtgtgttaacctggttataagggtgacggaa	1887
QY	2828	tcacatgacagctgtgtgactctctgcacaagaacaacggggctgtgcgaaggtgcgtcta	2887
Db	1888	tcacatgacacagctgtgtgattctctgcacaagaacaacggggctgtgcgaaggtgcgtcca	1947
QY	2888	agtgctcccaagaagaagcaccagttctctgtgcagctgtgcagaagaagcttacaagggtgatg	2947
Db	1948	gatctcccaagaagaagcagaaggtctctccctgcagctgtgcagaagaaggtatacaagggtgacg	2007
QY	2948	gctacagctgtcatgatagaataacccctgtgtcagaacaggtgtcaacgggggagatgcacatgagc	3007
Db	2008	ggcacagctgtcacagagataagataccctctgtcagagcgccttaacgggggtgtgtcacagagc	2067
QY	3008	acgcacactgtcaggaatgtaacgggtcccaagcaagataagtgttgaatgttaaaagtccaatg	3067
Db	2068	acgcacactgtgaagaatgatacagagcccggtgcaagcacaagtgltgaagttaaaagttcaatgt	2127
QY	3068	tcggggagcaggggtgagctgtgtagcctctgaacagctgtgcgcctgcagacgctgtgttacaagaca	3127
Db	2128	tcggggagcaggggtgaaactgtgtagcgcggagcagctgtgcacattgacacgctgtcttacaagaca	2187
QY	3128	acggacagctgcacaccacaatgtccagctgtgtgcagacaccttacttccagagacaagacgtag	3187
Db	2188	atggcagctgtcatalgcatagcagcccaatgtgtgtgcagccctccacttccagatataccactgttg	2247
QY	3188	gagatattccatctacgtcctccctctgtggccaagtatacaactgacatttgacaagaagccaag	3247
Db	2248	gggtgttccatctacgtcctccctctgtggccaagtatacaactgacatttgacaagaagccaag	2307
QY	3248	aagcctgtgcacaagaagaagctgcagacatlagcacacttacaacacagctctccatgtccaga	3307
Db	2308	aggcctgtgcacaagaagaagctgcagacatlagcacacttacaacacagctctccatgtccaga	2367
QY	3308	aggccaagtatcaacctgtgtcgg	3331
Db	2368	agacctgtatctcttaccacaag	2391
RESULT 5			
AAf87117			
ID	AAf87117 standard; DNA: 2512 BP.		
XX			
AC	AAf87117;		
XX			
DT	26-MAR-2002 (first entry)		
XX			
DE	NOV6 coding sequence.		
XX			
KM	NOV: cytosolic; contrareceptive; antiinflammatory; immunomodulatory;		
KM	cardiovascular; casein kinase II phosphorylation site; contrareption;		
KM	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;		
KM	epidermal growth factor; cell development; apoptosis; cell adhesion;		
KM	growth migration; cell structure; motility; cancer; immune disorder;		
KM	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;		
KM	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;		
KM	therapy. NOV1: NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;		
KM	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.		
XX			
OS	Unidentified.		
XX			
PH	Key		
FT	Location/Qualifiers		
FT	CDS 416..2425		
FT	/*tag= a		
FT	/product= "NOV6"		
XX			

XX MW020136638-A2.  
PD 25-MAY-2001.  
XX  
PF 17-NOV-2000; 2000OWO-US31543.  
XX  
PR 19-NOV-1999; 99US-16636P.  
PR 29-NOV-1999; 99US-167785P.  
PR 08-MAR-2000; 2000OUS-187844P.  
PR 16-NOV-2000; 2000OUS-0715417.  
XX  
PA (CURA-) CURAGEN CORP.  
PI Shlunkes RA, Lichenstein H, Vernet C, Fernandes E;  
XX  
DR WPJ: 2001-648134/74.  
DR P-PSDB; AAB83362.  
XX  
PT Novel human polypeptides and the nucleic acids that encode them useful  
PT for preventing, diagnosing and treating e.g. cancer, inflammation and  
PT immune disorders -  
XX  
PS Claim 9; Page 24-25; 141pp; English.

This sequence encodes the NOV6 protein. The invention relates to the NOV1-NOV16 proteins, and their coding sequences. The proteins have cytosolic; contraceptive; antiinflammatory; immunomodulatory; and cardiovascular activities. The sequences may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate NOVX expression. They may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of protein by expressing inactive proteins or to supplement the patients own production of protein. They are used to produce NOVX proteins, by inserting the nucleic acid into a cell and culturing it to express the protein. The DNA may be used as DNA probes in assays to detect and quantitate the presence of similar DNAs in samples, and which patients may need restorative therapy. The NOVX protein may also be used as antigens in the production of antibodies (Abs) against NOVX and in assays to identify modulators of NOVX expression and activity. The anti-NOVX Abs and antagonist are used to down regulate expression and activity. The anti-NOVX Abs are used for detecting the presence of NOVX in samples. Disorders that may be prevented, diagnosed and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5, NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites characteristic of serine/threonine kinases, and are used to treat kinase-related disorders (e.g. Pentz-Jeghers syndrome, cellular proliferation and contraceptive). NOV2-3, NOV6 and NOV8 are homologous to the epidermal growth factor (EGF)-like super family and are involved in, e.g. regulation of cell development, apoptosis, cell adhesion, growth migration, cell structure and motility and protein management, and are used to treat cancers, inflammatory disorders, immune disorders and cellular adhesion disorders. NOV6-10 are homologous to EGF-like fibillin proteins and are used to treat cardiovascular disease e.g. hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.

Sequence 2512 BP; 623 A; 657 C; 651 G; 581 T; 0 other:

```

Query Match          26.4%   Score 1243.2:   DB 23: Length 2512:
Best Local Similarity 80.5%:   Pred. No. 0:
Matches 1468; Conservative    0; Mismatches 353; Indels    3; Gaps    1
```

1508 cctcaatcatgagcccgagatgaagaactggatcagtcaggagcgtcatgcgccagtgc 1567  
||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 600 ccctgatcttcctgttaacagttaaagactggacaataaactgtttaaccgccagctc 659

1568 ttccgcatcacgttggtggcgtccagcagcgtcgttgttggacaacctaaagtggaccacaa 1627  
|||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 660 ttccgtacacatggtgctgcgtccaccacagctcctctggaaaaccctgtaattgatctcaa 719

1628 gggccacgacccctccaagagagccagtttcattctctctctcctaaggacagtcagtgtca 1687  
|||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 720 atgctactctcccctcaagagagccaatgcatctccgtctctctcctaagacagcgtgtata 779



QY	1668	taacaatgagcgagagtgctctgtccaagtgaatcaatcaagccaatggtcatalccag	1747
Db	780	taaacataaagcgaaagatcatatccagtgatatactactaactaagtgatgtttca	839
QY	1748	ttatagacaagtgtgtcttccccaanaacttgtrtaaccccnaaagtgctctggca	1807
Db	840	tcatagacaatgtctatctccccaanaaatttgcttatccctcccaagaacactctggaa	899
QY	1808	gggtcttgcnaaaatcttactatacagtgagcaagcaaacagagatatcaaatltagaagt	1867
Db	900	gaattctgcnaaaatcttagagcttcttgagcttgcaacaacaatggtctcatatcaaatltagcaact	959
QY	1868	tgatacagagctcaaggtctgtctcagltcaatcaatgaatccatccaaccccgatcag	1927
Db	960	taatacagagactcaggttcttgatgtgtatcaatccgatccatccacaccccgatcactc	1019
QY	1928	tctctgtccctacgagacaanaagcccttgaaagccttgcaccccaagcagcaggtccctgt	1967
Db	1020	tctctgtccctacccgacccaagccctcccaatgccttaccctgtgaaacaacaggtcttcgt	1079
QY	1968	tcaatlcaagagacaacaagagacaagctgaagtgcttacccttgaaagtccacagtcagagact	2047
Db	1080	tcaacccaagacaacaagagacaagctgaagagatatttgaaatttcatgtgatacgaagt	1139
QY	2048	ccaaggtcttgaacttgcagacctcccaaggtctgtctctctgtgaagacctgtcaaggtcag	2107
Db	1140	ccaaggtctttagcctgttgatcttcccaatcccatctgcctgtgaagacctgtcaaggtltag	1199
QY	2108	agctgagatgtgaagtgtgagagactgcagtgatactgtagagcttcttcttaaacagacaat	2167
Db	1200	agctgagatgtgaatgtgagagctgcagtgagacaatcagtgagaccttcttcgtaatgtccaa	1258
QY	2168	tgtagcaattcaatacacacccggagactctgttlltagctgggtgtgtagactltagatctgact	2227
Db	1260	ctgtcagaatgtgtgcagctgcagagctctgtgttgtaacctgggtgtgtgaccttagagctltagct	1319
QY	2228	gctcaatccatgaatccctacaccttgagtggtccgagatgtgaacacttctactactctgatattc	2287
Db	1320	gtctgcagatgtatgccacaccttggtgggtgcgcgtgtgacacacttctactactcttagatct	1379
QY	2288	cgggggagtgcgagaaagtcttgatcttcaactcccaatgacctggaagagacaagccaag	2347
Db	1380	cgggggagtggtggagagctgtgtcaataatctccagctgcaccaagtggtgaggtaaaccaag	1439
QY	2348	gcgtggaagaaagatgtatctataaacccgttactcttcaagagaaagtgtgaaggtctgc	2407
Db	1440	gtgtggaagcagaagtgtctctatacac--tgcccttcaagggaaacctgtgaaggtctgc	1496
QY	2408	agaacactgtgacccgttggtgatctccaacccccaagtggtgcagatgttactctatgcag	2467
Db	1497	ggggagcggtgcgaagcttggtgatatacaatccccaagtggtgcgaaggtctactctgcggcag	1556
QY	2468	actgtcagccctgcctgcggagagacagatatcaacccgtgtataaacccggagcatgtgcgg	2527
Db	1557	actgtcagccctgcctgcggagagacaggttgcgcccggtataaacccggggtgtctgcctgt	1616
QY	2528	atctgtatacaacccaatgagacagtgctctatgtgcacacccgggtcttcaacggacaagcttgc	2587
Db	1617	atcagtlactgcgcacacccggagagtgtaaatgtcaacaacccgggtcttcaatgtgagcgtgtg	1676
QY	2588	agctctgtctgcacgtgagagatttgggtgcctgcagctgtcatagcccccagctgcctcagagcatg	2647
Db	1677	agatgtgcttgcgcggggagatcttgggtccctgtatgtcttgcctgtgcctgtcatagacaag	1736
QY	2648	gacagctgtgatgagagatatacagagctccggggagagtgctctgtgtataaacagggtgagag	2707
Db	1737	gacagctgtgatgagatgagatcaacgggtcctccgggtgcagtgactctgtgtataaacgggtgagag	1786
QY	2708	cgcgtctgtgtgacacttcccaacagctgtatctgcaggtgtgtacaacactgtctgtcgtgc	2767
Db	1797	gcccctctgtgtgatacactcaagcaggttcttctgcaggtgtgtatacgcctctctgtctgtctc	1856

QY	2766	acgccacctgtacagagaacaacacacgtgtgtgtgttaacttgaacttaacgaaggtgacggga	2827
Db	1857	atcgacacctgtaaagagagaacaacacacgtgtgtgtgttaacttgaacttataaaggttaccgaa	1916
QY	2828	tcaactgaacagctgttgactcttcgcaacagaaacagggggctgttcggaagtcgcta	2887
Db	1917	tcaactgaacagctgttgactcttcgcaacagaaacagggggctgttcggaagtcgcca	1976
QY	2888	agttgtccccaagaagagaccacacagctcttcgcaagctgcaagaagaagcttacaaggggagt	2947
Db	1977	gattgtccccaagaagagagacgaaggtctcttcgcaagctgcaagaagggatacaagaaggagcg	2036
QY	2948	gctacagctgtcatagatagataccccctgttcgacacggtgtcatcaggggagatgcataagc	3007
Db	2037	ggcacagctgtcatagatagataccccctgttcgacacggtgtcatcaggggagatgcataagc	2096
QY	3008	acggcacctgtcagggatgacggggcccaagacgaactaaagtgtgaattgataaagttcaactatg	3067
Db	2097	acggcacctgtcagggatgacggggcccaagacgaactaaagtgtgaattgataaagttcaactatg	2156
QY	3068	tcggggagacggagtgtgactgtgaacctgtgacagctgtgcgctcagccgtgtgcttacaagaca	3127
Db	2157	tcggagatgtggctgtgactgtgacggtgtgacggagcagctgtcccatgtgacgcctgttacaagaca	2216
QY	3128	acggacagtgccacccacagatgcagctgttcgacacaccttacttccagagacagacgctag	3187
Db	2217	atgggcagtgccatgtcagacgacgaacatgtgtgcacaccttacttccagatgataccacgtgtg	2276
QY	3188	gagatttccatcttaagctcccccactctgggacgagtacaaacttgacatttgacaagaagccaaag	3247
Db	2277	gggtgttccatcttaagctcccccactctgggacgagtacaaacttgacatttgacaagaagccaaag	2336
QY	3248	aagcctgtgcacaagaacgtcgcagaccatagaccacttaaacacagctcttccatgtccagga	3307
Db	2337	agcctctgtgcacaagaacgtcgcagaccatagaccacttaaacacagctcttccatgtccagga	2396
QY	3308	aggccaagtatcaacctgtgtctcgg	3331
Db	2397	agacctgtattcttccattaccagag	2420
RESULT 6			
ID	AAf87113 standard; DNA: 2011 BP.		
AC	AAf87113:		
XX	26-MAR-2002 (first entry)		
DT	NOV2 coding sequence.		
DE	NOV2 coding sequence.		
XX	NOV: Cytostatic; casein kinase II phosphorylation site; contrareception;		
KW	cardiovascular; catenin kinase II phosphorylation site; contrareception;		
KW	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;		
KW	epidermal growth factor; cell development; apoptosis; cell adhesion;		
KW	growth migration; cell structure; motility; cancer; immune disorder;		
KW	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;		
KW	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;		
KW	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;		
KW	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.		
XX	Unidentified.		
XX	OS		
XX	FH		
FT	Key		
FT	CDS		
FT	Location/Qualifiers		
FT	877..1824		
FT	/*tag= a		
XX	/product= "NOV2"		
XX	WO200136638-A2.		
PD	25-MAY-2001.		
XX	17-NOV-2000; 2000WO-US31543.		

XX 19-NOV-1999; 9905-166336P.  
PR 29-NOV-1999; 9905-167785P.  
PR 08-MAR-2000; 2000US-187844P.  
PR 16-NOV-2000; 2000US-0715417.  
XX (CURA-) CURAGEN CORP.  
XX Shinkets RA, Lichenstein H, Vernet C, Fernandes E;  
PI MPI: 2001-648134/74.  
XX P-PSDB; AAB83358.  
DR  
XX Novel human polypeptides and the nucleic acids that encode them useful  
PT for preventing, diagnosing and treating e.g. cancer, inflammation and  
PT immune disorders -  
XX  
XX Claim 9; Page 10-13; 141pp; English.  
XX  
CC This sequence encodes the NOV2 protein. The invention relates to  
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have  
CC cytosolic; contraceptive; antiinflammatory; immunomodulatory; and  
CC cardiovascular activities. The sequences may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate NOVX  
CC expression. They may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of protein by expressing inactive proteins or to  
CC supplement the patients own production of protein. They are used to  
CC produce NOVX proteins, by inserting the nucleic acid into a cell and  
CC culturing it to express the protein. The DNA may be used as DNA probes in  
CC assays to detect and quantitate the presence of similar DNAs in samples,  
CC and which patients may need restorative therapy. The NOVX protein may  
CC also be used as antigens in the production of antibodies (Abs) against  
CC NOVX and in assays to identify modulators of NOVX expression and  
CC activity. The anti-NOVX Abs and antagonist are used to down regulate the  
CC expression and activity. The anti-NOVX Abs are used for detecting the  
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed  
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,  
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites  
CC characteristic of serine/threonine kinases, and are used to treat  
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular  
CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to  
CC the epidermal growth factor (EGF)-like super family and are involved in,  
CC e.g. regulation of cell development, apoptosis, cell adhesion, growth  
CC migration, cell structure and motility and protein management, and are  
CC used to treat cancers, inflammatory disorders, immune disorders and  
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like  
CC fibrillin proteins and are used to treat cardiovascular disease e.g.  
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.  
CC  
XX Sequence 2011 BP; 458 A; 551 C; 583 G; 419 T; 0 other:  
SQ

Query Match 25.4%; Score 1196; DB 23; Length 2011;  
Best Local Similarity 77.6%; Pred. No. 0; Mismatches 330; Indels 110; Gaps 2;  
Matches 1524; Conservative 0;

QY 2108 agctgaagtgtgaagtggtggaactggcagtgacatcgtgtgaagcttcttaacgaacaa 2167  
DB 398 agctgaagtgtgaagtggtggaactggcagtgacatcgtgtgaagcttcttaacgaacaa 457  
QY 2168 tgtcgaactatcaacacccggagactctgttctgaagtggtgtggtcctatgcatgact 2227  
DB 458 cctgcagaattgtgcagccggagactctgttctgaagtggtgtggtcctatgcatgact 517  
QY 2228 gactactatgaatcctaccagtggtgcagatgtgaactttactacttgatattc 2287  
DB 518 gtctgcagtattgacacccctgtgggcgtgtgtgacacttactacttgatattc 577  
QY 2288 cgggggagtgctggaagtgtgatttcacttcccaatgtcccaactgaagcaagcaag 2347  
DB 578 cgggggagtggtggaagtgtgttaatactccacgtgtcccaagtgtagtcaaccaaag 637  
QY 2348 gctgtgaagaagtgtatatacaacccgttactcttcaagaggaaagtggaagctgccc 2407  
DB 638 gtgtgaagaagtggtgtctcttaacac---tgccctcaagaggaaacttggaagctgccc 694  
QY 2408 agaacctgtgacccgtgtgtatccaaaccccaagtgctgcatgttactatgcacag 2467  
DB 695 gggagcagatgcagcctgtgtatatacagatcccaagtgctgcaaggtactctggcgag 754  
QY 2468 actgtcagctgtgcccctggagggagaccagatcacccgtgttaacaaacgggcatgtgccc 2527  
DB 755 actgtcagctgtgcccctggagggagaccagatgtacccgtgttaacaaacggggtgtcgtccttg 814  
QY 2528 atcgtatcaacaccatgtggaagtgacctatgtaccacccggttcaacgggagacagctgcg 2587  
DB 815 atcagtagtctgcgcagcggaagatttggcctgatttccgcctgtggtcgtccaaacacag 874  
QY 2588 agctcgtctggaatgtggaagatttggcctgacttcaagcccgagctgtcctcgaagatg 2647  
DB 875 agatgtcgtgcgcgggaagatttggcctgatttccgcctgtggtcgtccaaacacag 934  
QY 2648 gacagtgtaggaaggatcacagctccggggtgctccttctgtgaacaaagtgtagcag 2707  
DB 935 gacagtgtaggaaggatcacagctccggggtgctccttctgtgaacaaagtgtagcag 994  
QY 2708 ccgctcgtgtggaactcccaacagctgtatcgcagtgtagcaccctgtgtcctcgtgc 2767  
DB 995 gccctcgtgtggaactcccaacagctgtatcgcagtgtagcaccctccttctcgtgc 1054  
QY 2768 agccacactgtgagagaacaacacgtgtgtgttaacttgaactaagaagtgtagcagga 2827  
DB 1055 atgcacactgtgagagaacaacacgtgtgtgttaacttgaactaagaagtgtagcagga 1114  
QY 2828 tcaatgtgacagtgtagtgaacttctgcaaacagaaacgggggtgtgtcgaagtgtagta 2887  
DB 1115 tcaatgtgacagtgtagtgaacttctgcaaacagaaacgggggtgtgtcgaagtgtagta 1174  
QY 2888 agtgcctccagaagaagcaccacaaagctcttgcagctgtgcaagaagaagctacaaggagatg 2947  
DB 1175 gatgtccctcagaagaagcaccacaaagctcttgcagctgtgcaagaagaagctacaaggagatg 1234  
QY 2948 gctacagctgtcatagagataaacctctgtgcagacgtgtgtaaacgggggtgtgcatagagc 3007  
DB 1235 ggcacagctgtcatagagataaacctctgtgcagacgtgtgtaaacgggggtgtgcatagagc 1294  
QY 3008 agccacactgtgagataagacgggcccgaagcaagcaatgaatgtgaatgtlaaagctactatg 3067  
DB 1295 agccacactgtgagataagacgggcccgaagcaagcaatgtgtgaatgtlaaagctactatg 1354  
QY 3068 tccggagacgtgtgagctgtgagcctgtgacagctgtgcctgcagccgtgtcttaacagaga 3127  
DB 1355 tccggagacgtgtgagctgtgagcctgtgacagctgtgcctgcagccgtgtcttaacagaga 1414  
QY 3128 acggacagtgccacccagaatgcagctgtgtcagacacttacttccaggaacagacagctag 3187  
DB 1415 atggcagtgcatgtcagacagcacaatgtgtgcagacttccacttccaggaatcacactgttg 1474  
QY 3188 gagtattccatctaagctcccccactgggcagttacaacatgacatttgaacaaagcacaag 3247

```

Db 1475 ggggttcacatcagctccaccagctggccagctataagctgaccttgcacaagccagag 1534
Oy 3248 aagcctgtgccaaagaagctgacacatagcaccatacaacagctctctctgtccaga 3307
Db 1535 aggcctgtgccaaagaagctgacacataggcaacctcaacagcctctctatgtccaga 1594
Oy 3308 aggcacaatatacctgtgctcgccgagctgctgtagagatggcgagctgtctctaccga 3367
Db 1595 aggcacaatatacctgtgctcgccgagctgctgtagagacccggcgagctgtctaccga 1654
Oy 3368 ctgagctatgctctcagaagctgtgtagcaaacgtgtgtggagctgtagactacgatalcca 3427
Db 1655 cagcctgtgctccaccagactgtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1714
Oy 3428 gggccaacaagaatgtagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3487
Db 1715 gaccacaacaagaatgtagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1762
Oy 3488 cctgcaagcgagctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3547
Db 1763 ----- 1762
Oy 3548 tcatgtcctccctcgtctcacaacctctcagacagaagctgtgtgtgtgtgtgtgtgtgtgt 3607
Db 1763 ----- gaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1787
Oy 3608 cagcccgagcagcagcatttctgaaacacctgacctgacctgacctgacctgacctgacctg 3667
Db 1788 cagctcgagcagcagcatttctgaaacacctgacctgacctgacctgacctgacctgacctgt 1847
Oy 3668 ctgtgccacaagaacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3727
Db 1848 ctgtgccacaagaacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1907
Oy 3728 accacctacctaagtgtcaagctctctcttatacagcctgtgtcaatgtgtgtgtgtgtgt 3787
Db 1908 accacctacctaagtgtcaagctctctcttatacagcctgtgtcaatgtgtgtgtgtgtgt 1967
Oy 3788 ggaactatgtgtggaagcacaactgctcattcattcagccagagac 3831
Db 1968 aaacgagcgtgtgtggaagcacaactgctcattcattcagccagagac 2011

RESULT 7
AAF87114
ID AAF87114 standard; DNA; 1804 BP.
XX
AC AAF87114:
XX
DT 26-MAR-2002 (first entry)
XX
DE NOV3 coding sequence.
XX
XX NOV, Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
KM cardiovascular; casein kinase II phosphorylation site; contraception;
KM serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
KM epidermal growth factor; cell development; apoptosis; cell adhesion;
KM growth migration; cell structure; motility; cancer; immune disorder;
KM inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
KM cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
KM therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
KM NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.
XX
OS Unidentified.
XX
XX
XX Key Location/Qualifiers
FH 877..1611
FT CDS /*tag= a
FT /product= "NOV3"
XX
XX W0200136638-A2.

```

```

PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31543.
XX
PR 19-NOV-1999; 99US-166336P.
XX 29-NOV-1999; 99US-167785P.
PR 08-MAR-2000; 2000US-187844P.
PR 16-NOV-2000; 2000US-0715417.
XX
PA (CJPA-) CUBAGEN CORP.
XX
PI Shinkels RA, Lichenstein H, Vernet C, Fernandes E;
XX
DR WPI: 2001-648134/74.
XX P-PSDB: AAB93359.
XX
PT Novel human polypeptides and the nucleic acids that encode them useful
PT for preventing, diagnosing and treating e.g. cancer, inflammation and
PT immune disorders -
XX
PS Claim 9; Page 14-17; 141pp; English.
XX
XX This sequence encodes the NOV3 protein. The invention relates to
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC cytosolic; contraceptive; antiinflammatory; immunomodulatory; and
CC cardiovascular activities. The sequences may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NOVX
CC expression. They may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of protein by expressing inactive proteins or to
CC supplement the patients own production of protein. They are used to
CC produce NOVX proteins, by inserting the nucleic acid into a cell and
CC culturing it to express the protein. The DNA may be used as DNA probes in
CC assays to detect and quantitate the presence of similar DNAs in samples,
CC and which patients may need restorative therapy. The NOVX protein may
CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonist are used for down regulate
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g., regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovascular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX
SQ Sequence 1804 BP; 426 A; 487 C; 521 G; 370 T; 0 other:
XX
Query Match 21.6%; Score 1016.4; DB 23; Length 1804;
Best Local Similarity 82.0%; Pred. No. 8.1e-278;
Matches 1183; Conservative 0; Mismatches 256; Indels 3; Gaps 1;
Oy 1868 tgatacagacacagcctgtgctgtatcatcagcactcaccacccagcagctg 1927
Db 158 tgttcagacacagcctgtgctgtatcatcagcactcaccacccagcagctg 217
Oy 1928 tctctgtgctcagcagcagcctgtgagagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1987
Db 218 tctctgtgctcagcagcagcctgtgagagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 277
Oy 1988 tcaatcacaagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2047
Db 278 tcaatcacaagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 337
Oy 2048 ccaagccttagctcagacactcccaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2107

```

```

Db      338 ccaagttttagctgtggaatctcccaacatccacatcctcctggaagaccctgcaagttcag 397
Qy      2108 agcttgatgagtggtggaagcagtcagatcagtcagtcctctctctaaagaaacaa 2167
Db      398 agcttgatgagtggtggaagcagtcagatcagtcagtcctctctctaaagaaacaa 457
Qy      2168 tgtcagatcatalacacccggggaacctgtgttgagtggtggtggtcgtatgcatgact 2227
Db      458 cctgcagaattgtgcagccggagcctctgtttgacctgggtgtggtcctacgcatgtact 517
Qy      2228 gactatcatgaattctaccatagtggtggtcgtatgcatgacttactactctgatatctc 2287
Db      518 gtctgtcatgattgataccacccctggggccgctgtgacaccttaactactcgaatgct 577
Qy      2288 cggggagagtcggaagtgtgacttctcccaatgcccacatgaaagcaagcaag 2347
Db      578 cggggagagtcggaagtgtgacttctcccaatgcccacatgaaagcaagcaag 637
Qy      2348 gctggaagaagaagtgatctacacccgttacccttcagaaggaacgtgtgaaggtgcc 2407
Db      638 gtgtgaaagcaagtggtctctacaccc--tgccctcaagaagaaacctggaaggtgcc 694
Qy      2408 agaacctgtgacacgtgtgataccaaaccccaaggtgtcgtcgtatgactatgacag 2467
Db      695 ggagagcatgcagctgtgtatatacagatcccaaggtgtcgtcgtatgactatgacag 754
Qy      2468 acgtgacgctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 2527
Db      755 acgtgacgctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 814
Qy      2528 atctgtacacacccatggaagcagtgatctacacccgtcgtcgtcgtcgtcgtcgtcgtc 2587
Db      815 atcagtaactcggccacccggagagtgatgataacacccgtcgtcgtcgtcgtcgtcgtc 874
Qy      2588 agctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 2647
Db      875 agatgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 934
Qy      2648 gacagctgtgagagggagatcacaagctccggggagtgctcgtcgtcgtcgtcgtcgtcgtc 2707
Db      935 gacagctgtgagagggagatcacaagctccggggagtgctcgtcgtcgtcgtcgtcgtcgtc 994
Qy      2708 ccgtcgtgtgtgacacatcccaagcgtataltcagcagtggtgacacacgtcgtcgtcgtc 2767
Db      995 gccctcgtgtgacacatcccaagcgtataltcagcagtggtgacacacgtcgtcgtcgtc 1054
Qy      2768 accgcaactgtgacagagaacacacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2827
Db      1055 atgcacactgtgaaagagaacacacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1114
Qy      2828 tcacatgacacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2887
Db      1115 tcacatgacacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1174
Qy      2888 agtgcctccagaagaagcaacacgctcctgtcagcgtcgaagaagagtcacaaagggagtg 2947
Db      1175 gatgcctccagaagaagcaacacgctcctgtcagcgtcgaagaagagtcacaaagggagtg 1234
Qy      2948 gctacagctgtatagagatagacccctgtgacagcgtgtgaacggggagtgacagc 3007
Db      1235 ggcacagctgtacagagatagacccctgtgacagcgtgtgaacggggagtgacagc 1294
Qy      3008 accgcaactgtgacagagtgacgggcccagcaagcataagtgatgaagtaagaagtcactatg 3067
Db      1295 accgcaactgtgacagagtgacgggcccagcaagcataagtgatgaagtaagaagtcactatg 1354
Qy      3068 tccggggacggagtgagactgtgagcctgtgacgagctgtcgcgtcgcgtcgcgtcgtcgtcgtc 3127
Db      1355 tccggggacggagtgagactgtgagcctgtgacgagctgtcgcgtcgcgtcgcgtcgtcgtcgtc 1414
Qy      3128 acggaacatgtcacccaatgtccagctgtgacagccttactccagagacagacgctag 3187

```

```

Db      1415 atggagcgtgcatcgtcagacgccaatgtgtgacctccactccagatataccactgtg 1474
Qy      3188 gagtattccatcraagctcccaactgtggcaggtacaaactgcaattgtaaaagccaag 3247
Db      1475 ggggtgtctacatcactcccaactgtggcaggtacaaactgcaattgtaaaagccaag 1534
Qy      3248 aagcctgtgccaagaagcgtgcagccaatgcacactcaacacgctcctctatgcccaga 3307
Db      1535 aggcctgtgccaagaagcgtgcagccaatgcacactcaacacgctcctctatgcccaga 1594
Qy      3308 ag 3309
Db      1595 ag 1596

RESULT      8
AAAS7365
ID      AAAS7365 standard; DNA; 1259 BP.
AC      AAAS7365;
XX      03-OCT-2000 (first entry)
XX      DNA encoding a human hyaluronan-binding protein, designated BM-HABP.
DE      XX
XX      Hyaluronan-binding protein WF-HABP; OE-HABP; BM-HABP;
KW      Proliferative condition; metastasis; inflammation; ischemia;
KW      host defence dysfunction; immune surveillance dysfunction; arthritis;
KW      multiple sclerosis; autoimmunity; immune dysfunction; allergy; ss.
XX      Homo sapiens.
OS      Homo sapiens.
XX      Key
FH      Location/Qualifiers
FT      199..1257
FT      CDS
FT      /*tag= a
FT      /product= "hyaluronan-binding protein"
FT      /transl_except= (pos: 478..480, aa: Xaa)
FT      /transl_except= (pos: 667..669, aa: Xaa)
FT      /transl_except= (pos: 1105..1107, aa: Xaa)
FT      /transl_except= (pos: 1138..1140, aa: Xaa)
FT      /transl_except= (pos: 1156..1158, aa: Xaa)
FT      /transl_except= (pos: 1168..1170, aa: Xaa)
FT      /transl_except= (pos: 1171..1173, aa: Xaa)
FT      /transl_except= (pos: 1249..1251, aa: Xaa)
FT      /note= "Xaa is an unspecified amino acid; no
FT      termination codon given"
XX      MO200039166-A1.
XX      06-JUL-2000.
XX      20-DEC-1999; 99WO-US30462.
XX      23-DEC-1998; 98US-0113871.
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (AMNA-) AMERICAN NAT RED CROSS.
XX      Hastings GA, Liau G, Tsifrina E;
XX      WPI: 2000-452376/39.
XX      P-PSDB: AMY93913.
XX      New hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP,
XX      OE-HABP and BM-HABP, useful for treating proliferative conditions,
XX      metastasis, inflammation, ischemia, arthritis and multiple sclerosis -
XX      Claim 2; Fig 4A-B; 457pp; English.
XX      The present sequence encodes a hyaluronan-binding protein. The
XX      specification describes four hyaluronan-binding proteins, known as
XX      WF-HABP, WF-HABP, OE-HABP, and BM-HABP. The polypeptides are useful
XX      for treating diseases such as proliferative conditions, metastasis,

```

CC inflammation, ischemia, host defence dysfunction, immune surveillance  
CC dysfunction, arthritis, multiple sclerosis, autoimmunity, immune  
CC dysfunction and allergy.

Sequence 1259 BP; 295 A; 348 C; 346 G; 258 T; 12 other;

Query Match	18.8%	Score 883.4	DB 21	Length 1259
Best Local Similarity	81.5%	Pred. No. 4e+240		
Matches 1026; Conservative	2;	Mismatches 228;	Indels 3;	Gaps 1;

[illegible]

Db	901	gagaccaccctgcgaatgtcagcatgtttttctacaatgaacctgtcgaatgacacc	960
Qy	3784	ctgagagactatgctcggaaagcaactgtcattcaacttcagccagagacagctccacaa	3843
Db	961	ctgaaacagagcgtcgtggaagcaagctgcctcattcactgcagacagagaccctccaccg	1020
Qy	3844	---gagacaggtcttggatggaagatcattctcgagctggagacatcgcgcgcac	3900
Db	1021	acgagagccagcgtgtgatggaagagacactcgtgagatcctgtgcctccaat	1080
Qy	3901	ggaatccctcatatattcttgcgaacctttgagagtcctctcccaagcagcaagcgtcgc	3960
Db	1081	ggagtcacacatgtcattccagcgtctttaaagaacccctgcgcccggtgacttgncc	1140
Qy	3961	caactcgcctcggagcagagatlatctctgtccgcgcgtcgtgacatcgtgtcgtatgct	4020
Db	1141	caacctgntgtggagaaaggatctctctntgtacatccatcccggtgactcgtggtcgtgc	1200
Qy	4021	ctggagacttactcttacttcgcgtcgtctaaagcagcgaaccactggtctccagcgtttga	4079
Db	1201	ttgctgtcttactcctcacttcgtgataaacgcgaaacacatcgcgtctccancatttga	1259
RESULT	9		
AAAS7362			
ID	AAAS7362	standard; DNA; 6761 BP.	
XX	AAAS7362:		
AC			
XX			
DT	03-OCT-2000	(first entry)	
DE		DNA encoding a human hyaluronan-binding protein, designated WF-HABP.	
XX			
XX			
KW		Hyaluronan-binding protein WF-HABP; OE-HABP; BM-HABP;	
KW		proliferative condition; metastasis; inflammation; ischemia;	
KW		host defence dysfunction; immune surveillance dysfunction; arthritis;	
KW		multiple sclerosis; autoimmunity; immune dysfunction; allergy; ss.	
XX			
OS		Homo sapiens.	
XX			
XX			
FT	Key	Location/Qualifiers	
FT	CDS	170..6643	
FT		/*tag= a	
FT		/product= "hyaluronan-binding protein"	
FT		/transl_except= (pos: 6341..6343, aa: Xaa)	
FT		/transl_except= (pos: 6494..6496, aa: Xaa)	
FT		/transl_except= (pos: 6536..6538, aa: Xaa)	
FT		/note= "Xaa is an unspecified amino acid"	
XX			
PN	W0200039166-A1.		
XX			
PD	06-JUL-2000.		
XX			
FF	20-DEC-1999:	99WO-US30462.	
XX			
PR	23-DEC-1998:	98US-O113871.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(AMNA-) AMERICAN NAT RED CROSS.		
XX			
PI	Hastings GA, Liau G, Tsifrina E;		
XX			
DR	WPI: 2000-452376/39.		
DR	P-PSDB: AAY93910.		
XX			
PT	New hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP		
PT	OE-HABP and BM-HABP, useful for treating proliferative conditions,		
PT	metastasis, inflammation, ischemia, arthritis and multiple sclerosis -		
XX			
PS	Claim 2; Fig 1A-H; 457pp; English.		
XX			
XX			
CC	The present sequence encodes a hyaluronan-binding protein. The		
CC	specification describes four hyaluronan-binding protein, known as		

CC WF-HABP, WF-HABP, OE-HABP, and BM-HABP. The polypeptides are useful  
CC for treating diseases such as proliferative conditions, metastasis,  
CC inflammation, ischemia, host defence dysfunction, immune surveillance  
CC dysfunction, arthritis, multiple sclerosis, autoimmunity, immune  
CC dysfunction and allergy.

SQ Sequence 6761 BP; 1231 A; 2118 C; 2090 G; 1311 T; 11 other;

Query Match	14.8%;	Score 695.6;	DB 21;	Length 6761;
Best Local Similarity	51.4%;	Pred. No. 2.5e-186;		
Matches 2065;	Conservative	0;	Mismatches 1824;	Indels 125; Gaps 15;

QY	25	ctcgagacagatgcccacataattccattcttcaggagtacatlaattcaacaactgca	84
Db	2333	ctggacttggctgcctctcttcacgctcttcocggaaattgtctgacgacacatggttgg	2392
QY	85	agtgcatactgagtctcgcagatgcttatctgttctgtctgcacaacatgaagccatcgaa	144
Db	2393	cccacgattgagctctgcacatgctccacacacattttgtgtcccaacaacgcctccctggag	2452
QY	145	aactatatacaggagagagaagaagccacatctcttaagaagataattcagttacatgtg	204
Db	2453	gcc-----cagggcaacagcagatcacctctgaaagcagacacagttgcgcaccatgtg	2503
QY	205	gtctctggggaaaaagctctcttgaaatlgactgtcataacggcaatgcacogagagacatg	264
Db	2504	gtctctgggggaagcctcttcacatlgaaacccctcgggaagggtgagacaacgcaactcctc	2563
QY	265	ctggggtctctctaactctctctctctctctctctctcttcctgcgaatlgacagctgtatgtgaatgaa	324
Db	2564	ctgtgacctctgcacacatctgcatctctcttaaacacaagatggtccagacctgtggatgaacat	2623
QY	325	gtcccaataaactacacacaaatgtgcacatgataaaggagtgatccaatgtcttcggagaa	384
Db	2624	gtgcacctgtgaagagccccaatgctctgagggccctcgccgcctcgctgattgtctgtcgtggg	2683
QY	385	gtctctgaaattcaagaagaacagatgtgaaacataatgacacataattgtgagagggag	444
Db	2684	gtctctgagcgttgggtcccaagtcgctctgcctatagccaagctgagggctctgcggagaa	2743
QY	445	tgtggaagtgttcccaagcaagccctctgccactcgagacaaacacattagaagagc	504
Db	2744	tgtgtaaacatgcacacagaaattccgctctgacattcaacaggtctccagctgtgcagagacaccc	2803
QY	505	aggaatgtcatctatccatctactatcatatgagggaagagatccgtatcatctcgggtgcag	564
Db	2804	aggaatgagctgtgtcttaacgactctgctt-----ctcctctcccgggctgtct	2854
QY	565	ccacagtgctggaacaacacatcaataaagaagcctgtg-9ctgctctctttggcccacaa	623
Db	2855	tacacatgtgcacaaagaatccaagtgtgcgagactgtgcctctgttcttttggagcctg	2914
QY	624	tgcacaagctgcccgcgggagaggtccaatgtgtgtctctgtggaaacggtctctgtctgtgaa	683
Db	2915	tgtgagccacatgcacaaagggtgtctaaagggtgtgtgtctcaagccatgtggcaggtgcc-aggaa	2973
QY	684	cgtgtggaatggaactgcacatgcctgcacagtctgcggctgaggtctcaattggagacagcttga	743
Db	2974	caggtctctctgtgacagcgggagatgtgcacatgcacagaggtcttccatgaaagagcctgtga	3033
QY	744	aaactctgacatgagggagatgtatgtatccactctgagacaagaacatgactcttgtgtccatlg	803
Db	3034	ggtgtgtgagctgtgcccgcctcaacgggtcccaactgtgacacggaggtgtgtgactgtgcccattg	3093
QY	804	gagatgtlaagccaagagaccccttgggagacggtctccctgtgactgtgacgtctcggctggcag	863
Db	3094	gtctgtccagaggggtgtcgaagggtgacggaagtgtgtctgtlaacgtgtggtctgcaagg	3153
QY	864	agtgaagtgtgacatgtgagatccacacaagaacatctgcaacgggtccgttcaaccacatgc	923
Db	3154	ctctcgctctgtgacacagaaatcacccagcctctcagttcccttagagatgtgcagccccaatgc	3213

OY	924	caactgcctcttcgataccagacgcaaaagccttcgtagcaaaatggtcgagagatctccgagg	983
Db	3214	caactgccttcgataccagacgcaaaagccttcgtagcaaaatggtcgagagatctccgagg	3273
OY	984	gaatggaaacgctctgcacaaagccataatgcctgttagagacacagcaatltgaaagatgttctac	1043
Db	3274	caatgtgcaactctctgtctcgaaggtgaaacccctgcgcgcacagcgcaatgggggtctctccc	3333
OY	1044	aaagggcagctatgaataaagaacccacccacgaagaaacgggtgtgtgtgtgtgtgtgtgtgtgt	1103
Db	3334	tcattgcacaaactgttaacaaagtgtgcacccctgtgcagcgagacatgtcacactgtgcagagatgtcta	3393
OY	1104	taccgagacgacgcatcgtgtgtgccttgaataacccggtgtttgtggagaaacatgtgtgtct	1163
Db	3394	catgtgtgcacatcggtgagctgtgtgtgcagaanaatlaacagctgttccatccacccgggggtgtcgt	3453
OY	1164	tgcacagaatgtgcagatgtgcacacagacagaagccacacagggccgtctgttaactgtctgtcc	1223
Db	3454	ccaaatcaacgctcgtgagtgcatctcccaactgtgccccacagacagtgctctctgtcagctgtcgtga	3513
OY	1224	gaagtaacactgtgagatgtgaa--agctctgtcgtcgttatacaatgtctgtcctaacgaaaca	1280
Db	3514	gggtatacagcgggtatgtgcataccgacctgtgcagctcctgtgcacccctgtcttaagaacaa	3573
OY	1281	tgtgcgctgtgcagttccatttgcctctctgtgcacatacctgtgcagagatgtcaaaagatatctac	1340
Db	3574	tgtgagatgtgcagcccatatgtccacctgtgcacaaagacaaggtggatgtgcagagaaacatgttac	3633
OY	1341	ctgcacagcccaacatacac--gggtatgtgaatcgtctctccgggcgaacatctacacggga	1397
Db	3634	ctgcgcacacagcccaaccccgctgtgggtggacggtccctcaactgtcgtgtgcgcgagctgcgcgtgga	3693
OY	1398	gcttcccaagaaccccttgcagcttcccaagtaactcttccacagtgtgcagggscatgtcgtctcg	1457
Db	3694	gtctcttgagggataagcatgtccatcttcttcacgctctccgcctccctgtgatat-----aa	3747
OY	1458	agagcttgcgtgacccctgtgcccccttaacgcgtgtgtcgcccttgtcttaagctccctcaatca	1517
Db	3748	ggaagctcaagggcgaatgtggccttctccacacttccctgtgcgcagacagatactaaatgaagaa	3807
OY	1518	tgagcccccggatctaaagaactgtggaatcaagcagggccctcaatgtcccaagttcttcgtatca	1577
Db	3808	cctgtctgcaggaatgaagctgtgcccggatctcgtgtgcatactgcacagctcgtgtgtcttaaca	3867
OY	1578	cgttgtgtggtgtccagacagctgtcgttgtgtgcaacaactaaagtgcacacaagtgtgccaagac	1637
Db	3868	cgttgtgtggtgtcgtgcgcgtgtgcgggcagggacctgtgcagacaggggtacgccaagcgc	3927
OY	1638	cctccaagagagagccagatcttccatctctgtctctcagcagactgtgttcatataaacaatga	1697
Db	3928	cctctcaacaggacacacactgtcgttccagcagagaggggagccagacataactcaaatgactt	3987
OY	1698	ggcgaaggtctctgtccaggtatcatatcaagacaatgtgcgtcatccagttatagaaca	1757
Db	3988	cgcgcgcgtgtgtgtgcagcgaacatgaagccgtgtgaacgggtcatctcgtcatcttatgtacgc	4047
OY	1758	gttctgtcttcccaaaaactgttactacaccccaaaaagtgtcctgttggcagaggttctgtga	1817
Db	4048	gttctgtctgtcccccgagggcgtgtgcactgtggagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	4107
OY	1818	aaatcttactacagtgtgcagacaacaacaggaatatcccaaatccagaagtgtatcacaga	1877
Db	4108	aaatgttcaacgcgcgcgcgcacaggtctcgtgttataaagatcatcttaaggggtcctctgaagt	4167
OY	1878	ctcaagcttgcgttcaatcaatcagtctatccacaccccaagtaacgtcttctgtgc	1937
Db	4168	ggcgcgctcctgtccctgtctgtgagaggaatcccaataagcccttaacaatgtctgtgccc	4227
OY	1938	tgcggacaacacccctgtgaagcgtgtccccacagacagagacttctgttactacaaga	1997
Db	4228	caacagacgcgccttctcgagctcgtcctccggatgcagcagcgtgtgtgtatcatcagaaga	4287
OY	1998	caacaaaggacaaagctgtgaagtcttaactcgtgaagttccacgtgtatccggagatccaaagctt	2057

Db 4288 cccaccgtgacaagctagaacacatctcgcgggacacatgatccatgtccagagcctt 4347  
QY 2058 agcttaagacctccaccgagctcgtcttccttggaagaccttgaaagccacagagcgtatgt 2117  
Db 4348 ggcatactgacactgcccaccacttgagccacttcgaacatgacatggaagcccaactctctt 4407  
QY 2118 gaagtgtagaacttggcagtgacatcgtgtgagctcttcttaacgaacaaatgtgcaatt 2177  
Db 4408 ---ctccctgacagccgaaacgcggcccgctgagctcattggtgtagagatgatagtccgat 4464  
QY 2178 cataccagcgagactctgttttgaactgtggtgtgagcctatggaattgaactgtactactaat 2237  
Db 4465 tgtgcaacgagcaactgccccttgaggggtgagcctgacatgtgacatgacacagcgtctgga 4524  
QY 2238 gaatccatccactaggttgagcagatgttgacacttttaactaccttcgataatcccgggaggtg 2297  
Db 4525 gccacgtgcctctgtgtgtcgtcgtgtgacacacttgagacacgcgccttgcgactggaacac 4584  
QY 2298 cggagatltgcatl-----ttaactccaatgcccacttgaagagcaagcc 2342  
Db 4585 ctgcagcatctgtgtgagcctgagccacccctgtctcgagggttcaagagagcagggcagccc 4644  
QY 2343 aaagggc-----gtg 2352  
Db 4645 tgaagcctcgtgcgcgtcttcaaccgaaagtctgagcgtccctcgcgtgcactcttggg 4704  
QY 2353 aaagaagaagtgtactctcaaccggttaaccttcaaggagaaac-----gtcgaaag 2402  
Db 4705 attacgaagcgtctcgggtgccaccacacccactlttggttagagcccaagcctgagcaggg 4764  
QY 2403 ctgcgaagaacctgtgcacccgtgtgtgtatccaaacccagaggtgcgtcatgttactaat 2462  
Db 4765 ctgcgacagccaattgtgttcaaccacacacttgaaagcccaagcctgcgccttgtaactatgg 4824  
QY 2463 gccacgactgttcagagcctgccttgaggagccaagatacacgctgtatacaacgcgggagctgtg 2522  
Db 4825 cagtgtgtgcacaagcttgcgccttgaggccacagcagccctgtgtatgacgctgagcgtg 4884  
QY 2523 ccgggactctgtacacacccatlggagacagtgctatgtcacacacgcgcttcaacgggacagc 2582  
Db 4885 catgtgaagcagatgtgtgtgcagctggtgtgtcgtgtgcgttcaagcttctgtctgtaagaagc 4944  
QY 2583 ctgcgagactctgcgtgcatactggaagatttgggcgtgcactgtacagcccgcaactctccga 2642  
Db 4945 ctgtggaactctgtgtcctcgtgtgccttggcccatgttgaagccctgcgcgtgcactgt 5004  
QY 2643 gcatgtgacagtgtgtatgtgaagggtatcacagagctccggggaggtgcctctgttgaacaagggtg 2702  
Db 5005 gcatgtgcgcgtgtgtatgtgagggcctltggggggtctcgtgcctcgtcttctgtgtgaaggcgtg 5064  
QY 2703 gacagcgcgtctcgtgtgtacacatcccaacagctgtatctgcagatgtgtgacacactgtctgc 2762  
Db 5065 gactgtggccacgcgtgtgtgtgtgtacactgtgagcctltcacgcttgtgtatcccaacccctgtgc 5124  
QY 2763 cgtgtacagcaacccgtgtacgggaacaacagctgtgtgtgttaacttgaactgaactgaagttga 2822  
Db 5125 acccggaagcgtgtgtgcgtgtgcaggaacagctgtgtgtgtgcagccttggtcctatgtgaaggga 5184  
QY 2823 cgggatacacatgtgacagctgtgtgacactctgcacaaacagaaacagcggtgtgtgcgaagt 2882  
Db 5185 tggcgtgtgtgtatcagtgtagagacctgtgtccagagcgggtatgtgtcgtcagtgagca 5244  
QY 2883 cgtctaaagtgtcccaagaagaagcaccagaagtctctctgcagcgtgcagaagaagctacaagg 2942  
Db 5245 cggcaacctgttagccaggttaggaacaatagtgtacactgtcacccgctcgtcccgagctcaggg 5304  
QY 2943 ggtatgtctaacagctgtgcatagagatagaacccctgtgcagagcgtgttcaacagggagatgcc 3002  
Db 5305 tgtatgtgtgagactgtgcgcggccgaacccctgtcacagatgtgccaacgcgcgggctgtcag 5364  
QY 3003 ttagcagcgcacacatgtgacgggtccagggccaagcataagtgtgaatgtataaagtca 3062  
|||||

Db 5365 cgaagcagccaaactgtcttgagcaaccggccttgaaacacacagcgcgtgtgtatgtgcacgag 5424  
QY 3063 ctatgtcggggagcagtgtagactgtgagccctgagacagctgcgc-----tcgacgttctt 3119  
Db 5425 ctacgttaggcgagtgtactgtgcagtgtctgtgagaggttcggaaccacactgtgtgagcgtctt 5484  
QY 3120 acaagcaaacgagcagatgtgcacccacagatgcccagctgtgtgcaagccttacttactccagagac 3179  
Db 5485 gggccagccacagccgcctgcacatcagatgcatatgtacatgactctgtacatcttcacagagaa 5544  
QY 3180 gaccgttaggagatltcatatcagctctcccaacttggccaggtataacatgtacatttgaaca 3229  
Db 5545 acggcctgtgcgttcttcacacccacagccacgagccgtctatgttctgaacttlttcgga 5604  
QY 3240 agccaagaagcctgtgtccaaagaagcctgtgagaccatagccacccataacacagcctctcta 3299  
Db 5605 ggtcgtgagcgagcaatgtgcgaagcagagagccgtctctgtcttcaatctccctcagcctctgc 5664  
QY 3300 tgcacaagaagcccaaglatcacactgtgtcgcgcgcgtgtgcagagatgtgcggtgtgc 3359  
Db 5665 tgcacagcagctgtggtcttccacccctgctcatalgtgcgtgcgtgcgaatgtctccactgc 5724  
QY 3360 ctaccgcactacgtatgtcctctcagaaagtgtgtgcaaaagctgtgtgagatcgtatagcta 3419  
Db 5725 ccaccctgtgttctccctgtgtgcgagctgtgtgcaatgtgtcgtgtggcatagtcaagct 5784  
QY 3420 cggatccagggccaaacagatgtgaatgtgtgagatgtcttctgttaccggatagaagatgt 3479  
Db 5785 ggggtgcgcgcgaagaacactctcaagaacgtgtggaactacgtcttcgcgtgtgcgaagatgt 5844  
QY 3480 gaactgcacccgtgaagcagagctatgtgtgagatgtg---cttcgcgtgcagttggaacct 3536  
Db 5845 ggcctgcgcagatgcgcgaatgtgtctcgtgtgtgtagaaggtatcgaacgcgtgtgcaatgtgagact 5904  
QY 3537 gctgcaggtctcactatgtccttccctcgcgtctcaacaactctcgtacagaaggtgtcgtctt 3596  
Db 5905 gctgtatgtgtgtgtgcgtgcacatctctccacacttctatgtatgtatgtggctta 5964  
QY 3597 ttccaagagctcaagcccgagagacagcatltttgaacacactgtacttaacctgtccalcgc 3656  
Db 5965 tgcacaatgtccacccagcgggtgtcctgcagcttctcgtgacttccctgtgagtatgtgacgtta 6024  
QY 3657 tggcacctgttctgtgcacagcaacagtggtgtacccgggaataagagcgtgtgtgcgg 3716  
Db 6025 taagaacactctgtctccctgtgtcaatgtgaaggtcttgttgacaacatgtgcgctgagtgccc 6084  
QY 3717 ggcacttggagcacaacctcaactaattgtcaacgtctccttlttaacatgac-----ctt 3768  
Db 6085 aaacttggagctgcatgtccttcaacagcccaacctcttaagtgtcacaagcgcagccagggaa 6144  
QY 3769 gtcaatgtgtactcttctgagagactatgtgtggaagccaactgtcatcttaaccttaagccag 3828  
Db 6145 gttgtcttcggccacatcgaagcctcgaagcctcaatcatalcagtgtacgcaagccttgacaacag 6204  
QY 3829 gaaccagctccacaagagacca--gattgtgtgtagaagatccactctgcagctgtgagcat 3887  
Db 6205 ttctctggccctctgtgcccacaggtacagatgtgtgttagccggtatctaatctgtgtgagcat 6264  
QY 3888 catgcgcgaatgtgaatctcatalatattctgtgaacctttagagatctccc 3941  
Db 6265 catgtccctcaatgtcatatccatcagtctgtgcagagccctcctcgtgcaccccc 6318  
RESULT 10  
AAC76373 standard: cDNA: 1377 BP.  
XX  
AC AAC76373:  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORF1928 polynucleotide sequence SEQ ID NO: 3855.  
XX





Qy	4333	cgcacagagccaaaggagccacatgcttgcctgtatataccgaggggagactccacactccctctgagcc	4452
Db	1139	ttctctgctgggctgagatgctgttg-ctgtgccacaccagactacagcttctctcctctgaccc	1197
Qy	4453	tat	4455
Db	1198	tct	1200

```

RESULT 11
AAAF87118
ID      AAAF87118 standard; DNA; 1624 BP.

```

AA AF87118; AC

DT 26-MAR-2002 (first entry)

NOV7 coding sequence.

NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;

KW **cardiovascular; casein kinase II phosphorylation site; contraception; cellular proliferation; Peutz-Jeghers syndrome**

KW epidermal growth factor; cell development; apoptosis; cell adhesion; growth migration; cell structure-motility; cancer; immune disorder; KW

KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome; cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome

KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;  
KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; NOV17; NOV18; NOV19; NOV20;

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818 2

XX  
XX

FT	CDS	387.1532
----	-----	----------

FT /product= "NOV7"

PN WO200136638-A2.

PD 25-MAY-2001.

PF 17-NOV-2000; 2000WO-US31543.

19-NOV-1999; 99US-166336P.

PR	08-MAR-2000; 2000US-187844P.
EN	22 NOV 1333; 9903-107782P.

PR 16-NOV-2000; 2000005-0/1341/.  
XX XX

PA (CURA-) CURAGEN CORP.  
XX

PI Shimkets RA, Lichenstein H, Vernet C, Fernandes E,  
XX

DR WPT; 2001-648134/74.  
DB P-PCDB: AAP83363

XX  
DT

Memo] bu... col...

PT for preventing, diagnosing and treating e.g. cancer, inflammation and

XX  
XX

This sequence encodes the NOV7 protein. The invention relates to the NOV1-NOV16 proteins, and their coding sequences. The proteins have cytosolic; contriceptive; antiinflammatory; immunomodulatory; and cardiovascular activities. The sequences may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate NOVX expression. They may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of protein by expressing inactive proteins or to supplement the patients own production of protein. They are used to produce NOVX proteins, by inserting the nucleic acid into a cell and culturing it to express the protein. The DNA may be used as DNA probes in assays to detect and quantitate the presence of similar DNAs in samples, and which patients may need restorative therapy. The NOVX protein may

CC also be used as antigens in the production of antibodies (Abs) against  
CC NOVX and in assays to identify modulators of NOVX expression and  
CC activity. The anti-NOVX Abs and antagonists are used to down regulate  
CC expression and activity. The anti-NOVX Abs are used for detecting the  
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed  
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,  
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites  
CC characteristic of serine/threonine kinases, and are used to treat  
CC kinase-related disorders (e.g., Peltz-Jeghers syndrome, cellular  
CC proliferation and contrareception). NOV2-3, NOV6 and NOV8 are homologous to  
CC the epidermal growth factor (EGF)-like super family and are involved in,  
CC e.g., regulation of cell development, apoptosis, cell adhesion, growth  
CC migration, cell structure and motility and protein management, and are  
CC used to treat cancers, inflammatory disorders, immune disorders and  
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like  
CC fibrillin proteins and are used to treat cardiovascular disease e.g.  
CC hypertrophic cardiomyopathy, Long-QT syndrome and marfan syndrome.  
XX

Sequence 1624 BP; 417 A; 420 C; 386 G; 401 T; 0 other;

Query Match	13.5%	Score 637.6;	DB 23;	Length 1624;
Best Local Similarity	76.0%	Pred. No. 3.3e-170;		
Matches 830; Conservative	0;	Mismatches 224;	Indels 38;	Gaps 2

QY	1508	ccctcaatcatgagccccggagataaagaactggatcagcaggagcctcatgctccaggcttc	1567
Db	571	ccctctgctctctgcttactaagactctgtaaaagactctgtaacaaatacagctttaaagccccagtttc	630
QY	1568	ttgcgtcaacagtggtggagctgcgcagcagctgctgtgttgagacacacctaaagagtagaccaca	1627
Db	631	ttcggtacacatggtgtgcgcgtgcaccagctgcttcctgtaaaacctgtaattgatactcca	690
QY	1628	gtgcgcacagacctccccaagagagagccagcttccatctcctgtctctcaagagacactgtgtca	1687
Db	691	atgtaactctccctcccaagagagccaatagtaactcgtctctcagaagcagctgata	750
QY	1688	taaacatctggagcggaaggtccctgtccatgagacatacagaccaatggcgtatccacg	1747
Db	751	taaacataaagagctaaagatatactaccagatatactccagtaactaactggaattgttca	810
QY	1748	ttatagaacagctgtgcgtctctccccaanaactgtctatcaccccaagaagtcgtgggca	1807
Db	811	tcatagaacaattgtctatctcccaaaaatttgcttatctaccctcccaagaacaactctgaa	870
QY	1808	gggtctctgcacaanaactctactacacagtggcagccaacaacagatataccaattcagacaag	1867
Db	871	gaattctgcacaanaactctatgagcttctggccaacaacaagtgctataccaatttagcaact	930
QY	1868	tgatacagagactcaagctctgctgtcagtaactaatcagtaactccatccacacccagtaactg	1927
Db	931	taatacagagactcaagcttctgtgagtaactataccagatcccatccacacccagtaacttc	990
QY	1928	tctctctgcacctgaagaaagccctctgaaagcctctggcccacagacagacagacactcctgt	1987
Db	991	tctctctgcccacacagacacaaagccctcccatgcccctaacctgcgtgaacaacagacactccgt	1050
QY	1988	tcaatacagaacaacaagagacaaagctgtgaagtctctacccttgaaagtcttcacagtccgagact	2047
Db	1051	tcaaccaagaacaacaagagacaaagctggaagagatattgaaagttctatgtaagaagtg	1110
QY	2048	ccaaagcctttagcttcaagacctccccaagctctgctctccctggaagaacctgtcagaagctcag	2107
Db	1111	ccaaagctttagcctgtgtgaatctctccccaacatccacatctcctctggaagaacctgtcagaagctcag	1170
QY	2108	agctgagatgtgaagtggtggaacctgcagctgacagatcagtgagactctcttcaacggaacaa	2167
Db	1171	agctgagatgtggaatgtgtgagcctgcggaaggacacatcgtgaaacctcttctgtaagtccaa	1230
QY	2168	tgtgagatctcatatacaccggagacctctgtttgacgtgggtgtggtctatgtgacattgact	2227
Db	1231	ccgtgcagaattgtgacagcggagagcctctgtttgacctgggtgtggtcctagagcattgact	1290

```
OY 2228 gacctactcaatgaatcctactacccctaggtggtccgagtgtgacactttactactcgtatctc 2287
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1291 gtctgcgtatgtatccaccctcgtggtggtccgctgttgacaccttactactcgtatgctc 1350
OY 2288 cggggggggtgcggaagtgtgacttctcactcccaatgccactgaaagagaagcacaagg 2347
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1351 cggggggggtgtggagctgtgtcaatactcccaactgcccagaagtgtgagtaaaccaagg 1410
OY 2348 gctgtagaagaagaatgtatctatacaccgttactcttcagagaagaagtgaaagctgccc 2407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1411 gtgtgagaagcagaagtgtctctacaacc---tgccctcaagagagaactgtgaagcgtgcc 1467
OY 2408 agaactgtgacacgggtgtgtatcccaaccccccagggtgtccatggttactctatgccag 2467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1468 gggggggtgtgcagccctgtatatacagatcccca----- 1500
OY 2468 acgttcagcgtgctgctgggggagaccagatataccggtgttaaacccggggcgtgtgccgg 2527
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1501 -----gctctgcctggggagaccagatgcccgtgttaaacccgggggtgtctgctctg 1552
OY 2528 atctgtacacacccatgtgacacagtgctatgcccacccggttcaacgggagacagcctgcg 2587
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1553 atcagctactcgtgcaccccgagagtgatgtaatgcaacacccggttcaatggtgacggcgtgtg 1612
OY 2588 agctcgtcgtgc 2599
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1613 agatgtgtgtgc 1624

RESULT 12
AAE87121
ID AAE87121 standard; DNA: 1577 BP.
XX
AC AAE87121:
XX
DT 26-MAR-2002 (first entry)
XX
DE NOV10 coding sequence.
XX
KM NOV: Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
KM cardiovascular; casein kinase II phosphorylation site; contraception;
KM serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
KM epidermal growth factor; cell development; apoptosis; cell adhesion;
KM growth migration; cell structure; motility; cancer; immune disorder;
KM inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
KM cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
KM therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
KM NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 387..1391
FT FT /*tag= a
FT FT /product= "NOV10"
XX
XX WO200136638-A2.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US31543.
XX
XX 19-NOV-1999; 99US-166336P.
XX 29-NOV-1999; 99US-167785P.
XX 08-MAR-2000; 2000US-187844P.
XX 16-NOV-2000; 2000US-0715417.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Lichenstein H, Vernet C, Fernandes E;
XX
XX WPI: 2001-648134/74.
XX
XX P-PSDB; AAB83366.
```

```
XX
PT Novel human polypeptides and the nucleic acids that encode them useful
PT for preventing, diagnosing and treating e.g. cancer, inflammation and
PT immune disorders -
XX
PS Claim 9; Page 35; 141pp; English.
XX
CC This sequence encodes the NOV10 protein. The invention relates to
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC cytosolic; contraceptive; antiinflammatory; immunomodulatory; and
CC cardiovascular activities. The sequences may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NOVX
CC expression. They may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of protein by expressing inactive proteins or to
CC supplement the patients own production of protein. They are used to
CC produce NOVX proteins, by inserting the nucleic acid into a cell and
CC culturing it to express the protein. The DNA may be used as DNA probes in
CC assays to detect and quantitate the presence of similar DNAs in samples,
CC and which patients may need restorative therapy. The NOVX protein may
CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonist are used to down regulate the
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovascular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX
SQ Sequence 1577 BP; 424 A; 410 C; 342 G; 401 T; 0 other;

Query Match 11.0%; Score 517.4; DB 23; Length 1577;
Best Local Similarity 78.8%; Pred. No. 4.e-136;
Matches 617; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

OY 1508 ccttcaatcagagcccgatlaaagactggagtcgacgagcgtcattgtccaggctc 1567
      ||| ||| | | ||||| ||||| | | | | ||||| ||||| ||||| |||||
Db 571 cctctgattcttctgttaccagtttaaagactggagcacaatacgtgttattgccaggctc 630
OY 1568 ttgcgtacacgtgtgtgtgcgcagcagctgtctgttggacaacctaaagtgcaccaa 1627
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 631 ttgcgtacacgtgtgtgtgcgcagcagctgtcttggaaacctaaagtgcaccaa 690
OY 1628 gtgcacagacctccaaggagagcagattccatcttcttccagacactgtgttca 1687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 691 atgctattcctccccaaggagagcacaatgcatcttcgttccagagacggtgtata 750
OY 1688 taacaatgagcgaaggtctcctgcagtgacatcatcagacacagcgtgtccacag 1747
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 taacaatgagcgaaggtctcctgcagtgacatcatcagacacagcgtgtccacata 810
OY 1748 ttatagaacaagtgtgtctcccaaaaactgtctatccaccccaagaatgtcgtgca 1807
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 tcatagacaatgtctatctcccaaaaattgtctatccactcccaagaacaactgtgaa 870
OY 1808 gggttctgcaaaaactctactacagtgagcagaaccccgatataccaatcagcaagt 1867
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 gaattctgcaaaaactctactacagtgagcagaaccccgatataccaatcagcaact 930
OY 1868 tgatcagagactcaggtgtgtgtcagatcatcagtgctcatcaccacccacagtcctg 1927
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 taatacagactcaggtgtgtgtcagatcatcagtgctcatcaccacccacagtcactc 990
```

OY	1928	tcctctgacctacgcagcaaaagccctggaaagccttgcgcccaagacagcagcattccigt	1987
Db	991	tcctctggccaccgcagcaccagccctccaaagccctactactctgacacacagcagcttcctgt	1050
OY	1988	tcaatcagaacacaaagcagcagcctggaattcttactactgaattccacgtgattccggact	2047
Db	1051	tcacccaagacacacaaagcagcagcctggaagagatattgaaattccatgtgatacgagatg	1110
OY	2048	ccaagcctttagcttccagaccctcccaaggtctgctctccggaagacccctgcgaagctcag	2107
Db	1111	ccaagcttttagctgtgtgacttctccacatccactgcgcggagaaccccgcaagtgctcg	1170
OY	2108	agcctgagtgctgaggtgtgtggaactgtgcagtgacatccgtgtgagctctttctaacgcaaaa	2167
Db	1171	agcttgatgtgaaatgtagtgagctgtggaagatcatcggtgactcttcttcgaatgacgaaa	1230
OY	2168	tgtgagatctcatacaccggagaccctgtgttgacgttggtgtgtgacctatgcatgtgact	2227
Db	1231	ccctgacagatctgtcagcgcggagccctctgttctgacctgggtgtgtgacctacgcatgtact	1290
OY	2228	gcctactcatgatactcactcaagctggtgcgagatgtgacacttcttactactgatatcc	2287
Db	1291	gtctgtcgtgattcatccaccctcggggggcgcgtgtgacaccccttactacttcgatgctt	1350
OY	2288	cgg	2290
Db	1351	cgg	1353
RESULT	13		
ID	AAA57363		
AC	AAA57363 standard; DNA; 1522 BP.		
XX	AAA57363;		
DT	03-OCT-2000 (first entry)		
DE	DNA encoding a human hyaluronan-binding protein, designated WF-HABP.		
XX			
KW	Hyaluronan-binding protein WF-HABP; OE-HABP; BM-HABP; proliferative condition; metastasis; inflammation; ischemia; host defence dysfunction; immune surveillance dysfunction; arthritis; multiple sclerosis; autoimmunity; immune dysfunction; allergy; ss.		
OS	Homo sapiens.		
XX			
EH	Key	Location/Qualifiers	
FT	CDS	31..1404	
FT		/*tag= a	
FT		/product= "hyaluronan-binding protein"	
FT		/transl_except= (pos: 1102..1104, aa: Xaa)	
FT		/transl_except= (pos: 1255..1257, aa: Xaa)	
FT		/transl_except= (pos: 1297..1299, aa: Xaa)	
FT		/note= "Xaa is an unspecified amino acid"	
XX			
PN	WO200039166-A1.		
XX			
PD	06-JUL-2000.		
XX			
PF	20-DEC-1999;	99WO-US30462.	
XX			
PR	23-DEC-1998;	98US-0113871.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(AMNA-) AMERICAN NAT RED CROSS.		
XX			
PI	Haslings GA, Liau G, Tsifrina E;		
XX			
DR	WPI: 2000-452376/39.		
XX	P-PSDB; AAY93911.		
PT	New hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP		
OE-HABP and BM-HABP, useful for treating proliferative conditions,			

XX	PT	metastasis, inflammation, ischemia, arthritis and multiple sclerosis -
XX	PS	Claim 2: Fig 2A-B: 457pp: English.
XX	XX	
CC	CC	The present sequence encodes a hyaluronan-binding protein. The
CC	CC	specification describes four hyaluronan-binding protein, known as
CC	CC	WF-HABP, WF-HABP, OE-HABP, and BM-HABP. The polypeptides are useful
CC	CC	for treating diseases such as proliferative conditions, metastasis,
CC	CC	inflammation, ischemia, host defence dysfunction, immune surveillance
CC	CC	dysfunction, arthritis, multiple sclerosis, autoimmunity, immune
CC	CC	dysfunction and allergy.
XX	XX	
SQ	Sequence 1522 BP; 277 A; 468 C; 454 G; 312 T; 11 other:	
Query Match	4.7%; Score 220.8; DB 21; Length 1522;	
Best Local Similarity	53.0%; Pred. No. 8.4e-52;	
Matches 571; Conservative	0; Mismatches 492; Indels 15; Gaps	4.;
0Y	2879 aggtcgtctaagtgatgcctcccaagaaggacccaagtctcttcgacgtcgaagaaggctaca	2938
Db	2 agcacagccacactgtatgcaggtatgaacaatgttcaattgtaccctgcgcctgcgcctacg	61
0Y	2939 aaggggatatgtctacagctgtcatagatagacccctgtgcacagcggtgtcaacgggggat	2998
Db	62 agggatgatgtctgcagcgctgcgcggccgaacccctgcacagatgtgcacacgcgggggct	121
0Y	2999 gccatagagacgcacactgtcagatagacggtgcgcgcgcgcgaagcataagttgataataa	3058
Db	122 gcaagcgaagcagccacactgtctgaagcacgcgccttgaacacacgcgcctgtgagtgccag	181
0Y	3059 gtcacatgtctggagagcagagtgagctgtc---gaagcttagagctgtgcgcgcgcacgtt	3115
Db	182 cagcgtatcagtgacgtatgtgacgtcagtgctgtctggaagatgcgaacacactgtgcagcgt	241
0Y	3116 gcttcacagagacaaagcagctgtccaccccaatgtccacagctgtgcacacactctaccag	3175
Db	242 gcttggtgcagcagccagccctgcgaactcagatgtgcattgtgacatgacctgtccacag	301
0Y	3176 acacgacgcgtatgagatattccattccatcagctcctccacactgggcaggtacaacatgcattg	3235
Db	302 aqaacagcgctggtcgcttltccacctccacccacacagcggcccttatgtgtcaacttlt	361
0Y	3236 acaaaagccaaaagaagcctgtgcacaagaagctgtgcacatagcacactacaacacgctc	3295
Db	362 cggagagctgaagcgcgatgtgcgaagcagaagggccgctcttctcatctccctcagctc	421
0Y	3296 cctatgtcccaagaagggccaagtatcacctgtgtgcgcgcgtgtgcgtgtggaagtgggcggg	3355
Db	422 ctgtgtcccaagcagctgtggcttccacacgtgtgcctcaatgggtcgtgtgcacatgtctca	481
0Y	3356 ttgctaccccaagctacgtatgcctctcagaagttgtgtgcacaaagttgttggatcgttag	3415
Db	482 ctgcacaccccttggcttltccctgtgtgcgcagctatgtgcattgtgcgggtgcacatgtca	541
0Y	3416 actacagatccaggggcacaaagagtgaaatgtggatgtcttctgttaccggatgaaag	3475
Db	542 gccgtgggtgcccgaagaacacctcagaagcgtggagatgtccactacgtctccgtgtgcaag	601
0Y	3476 atgttaactgtcacctgtcaagagcagctatgtgtggagatgtgcttct---cgltcagtgga	3532
Db	602 atgtgtgcctgcgcagatgtgcgaatgtcttctgtgtgacggtgtcagacgkgcaatgtgga	661
0Y	3533 acctgtcgcagttcctcatgtctctctccctcgcgtccacaacattcctgcacagaagtgtcg	3592
Db	662 agctctcgtgatltgtcgtgtgcacgtgcacaaattctccacctctcatatgtgagatgtatgtg	721
0Y	3593 cttttccaaagagcttcagcgcgcagagacaggaatttttgaacaactgtatgcctgtcca	3652
Db	722 gctatgtccaaatgtccaccacgcgggtgtctgcacttccctgtgaattccctgtatgttagctca	781
0Y	3653 tccgttgacaccctgttctgtccacagaacagtggtgtaccggaataaagaacgtctgt	3712

```
Db      782 cgtataagacactcttcgtccctgctcattgaagagcttctgtgacacacatgacgtgagtg 841
Oy      3713 gccggagacattgagcagccaccctcactaatgtaacgtctcccttctaagaac----- 3765
Db      842 gcccaaaccttgagcgtcgtatgctctcaaacgcccactcctaagtgcacaaagccagccag 901
Oy      3766 -cttgcaatgtgacacttcttgaggaactatgctgggaagcaactgtctacttaactcag 3824
Db      902 ggaagtgtctcgcgcgcacactcagcctcactcactcactcactcactcactcactcactcactc 961
Oy      3825 ccagagccagctccacccaagaacca-ggtttgtgattgtaagaatcattctgcagtggg 3883
Db      962 acagttcctctggccctcgttgcccccaggagacagttgtgttagtcgtatcattgtgtgg 1021
Oy      3884 acatctatgccgcgcgaatggaactcctcactatatttcttgaaactttgagagctccctcc 3941
Db      1022 acatcatgcttcaatgacatcattcattcactgtctgacagcccccctcctgtgaccccc 1079
```

## RESULT 14

```
AA500854
ID      AA500854 standard; cDNA; 718 BP.
```

```
AC      AAS00854;
```

```
DT      04-JUL-2001 (first entry)
```

```
XX      Human cDNA clone HLMIT84 encoding cancer related protein 28.
```

```
XX      Human; cancer related protein; HLMIT84; food additive;
XX      KW preservative; immunogen; antibody; bone cancer; adrenal cancer;
XX      KW bone marrow cancer; breast cancer; gastrointestinal cancer;
XX      KW liver cancer; lung cancer; urogenital cancer; immune disorder;
XX      KW Addison's disease; allergy; autoimmune haemolytic anaemia;
XX      KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX      KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX      KW acquired immunodeficiency syndrome; AIDS; cardiovascular disorder;
XX      KW myocardial ischaemia; wound healing; neurological disorder;
XX      KW Parkinson's disease; Alzheimer's disease; cerebral anoxia; epilepsy;
XX      KW viral infection; bacterial infection; fungal infection;
XX      KW parasitic infection; agonist; antagonist; ss.
```

```
XX      Homo sapiens.
```

```
OS      Key Location/Qualifiers
```

```
FT      CDS 1..321
```

```
FT      /product= "Cancer related protein 28"
```

```
FT      /partial
```

```
FT      /note= "No start codon"
```

```
XX      MO200118014-A1.
```

```
XX      PD 15-MAR-2001.
```

```
XX      PE 30-AUG-2000; 2000WO-US23794.
```

```
XX      PR 03-SEP-1999; 99US-0152296.
```

```
XX      PR 06-OCT-1999; 99US-0158003.
```

```
XX      PA (HUMA-) HUMAN GENOME SCI INC.
```

```
XX      PI Roschke V;
```

```
XX      DR WPI: 2001-235186/24.
```

```
XX      DR P-PSDB; AAU00893.
```

```
XX      PT Twenty nine nucleic acid molecules encoding human cancer associated
```

```
XX      PT proteins, useful in the prevention, treatment and diagnosis of cancer,
```

```
XX      PT immune disorders, cardiovascular disorders and neurological diseases -
```

```
XX      PS Disclosure: Page 383; 427pp; English.
```

```
CC      CC The sequence encodes a novel Human cancer related protein. The
CC      CC polynucleotides and polypeptides are useful for preventing,
CC      CC treating or ameliorating a medical condition in e.g. humans,
CC      CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC      CC polypeptides can also be used as a food additive or preservative to
CC      CC increase or decrease storage capabilities. The polynucleotide are
CC      CC useful for chromosome identification. The nucleic acids, protein,
CC      CC antibodies, agonists and antagonists are useful in the diagnosis,
CC      CC treatment and prevention of cancer (e.g. cancers of the adrenal gland,
CC      CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC      CC urogenital), immune disorders (e.g. Addison's disease, allergies,
CC      CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC      CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC      CC colitis), acquired immunodeficiency syndrome (AIDS), cardiovascular
CC      CC disorders such as myocardial ischaemias, wound healing, neurological
CC      CC diseases (e.g. Parkinson's disease, Alzheimer's disease, cerebral anoxia
CC      CC and epilepsy) and infectious diseases such as viral, bacterial, fungal
CC      CC and parasitic infections. Numerous examples of each type of disorder are
CC      CC given in the specification.
```

```
XX      SO Sequence 718 BP; 166 A; 206 C; 165 G; 181 T; 0 other;
```

```
Query Match 3.8%; Score 177.4; DB 22; Length 718;
```

```
Best Local Similarity 63.9%; Pred. No. 1.2e-39;
```

```
Matches 334; Conservative 0; Mismatches 181; Indels 8; Gaps 4;
```

```
Oy      3937 cctccacagcacaacagcgtgcgccactctgcccgtggagacgatattctgtccgtc 3996
```

```
Db      10 cccctgtcccccgtgaccttgaccacacacgtgctggagacgatctcttgcacatc 69
```

```
Oy      3997 gtctgtgtaactgtgctgattgtctgtgacgtctactcttaactcgcgcaagcagca 4056
```

```
Db      70 atctcgtgactgtgggtgtgtgtctgtgtcctgtactccactccttggaataaccggaga 129
```

```
Oy      4057 accactggttccagcggttttga-tcagaagagacattgattgtcgtcttggcaag 4115
```

```
Db      130 acaatcgcttcacgaatlttgagtcggaagaggaacttaatgttgacgtccctggagag 189
```

```
Oy      4116 cagaagcccaagaatctcgcaaacacctctgtatgagacctcagccgcgcaacccacagag 4175
```

```
Db      190 cagagccttgagaatctatctgaacctctgtatgagacacacactcagctcccccagaa 249
```

```
Oy      4176 tctctcgtgtgaccccttcacagaccctggagacagatctggagagcagggaccctctg 4225
```

```
Db      250 ccttcctcagcacccttcacagacctctgaagaaaggagcttgagggaatgacaccccttg 309
```

```
Oy      4236 ggggcactg--cgctctgacatgagaagcagcagcaacacacacatcacggttccag 4292
```

```
Db      310 aggaactgtgagggccttgagcgggagatgacacatcactcactcctggccat 369
```

```
Oy      4293 gtgattccagcccccagctgtctcatgattcagttgttttaagaatgacaacactcata 4352
```

```
Db      370 caactgtgaattctcagc--accagttgcctttagaagacgtaaatguccttaagact 426
```

```
Oy      4353 agccagccatacctcaccctctgtttaattgttgatgttcgcaaggcctaaggagact 4412
```

```
Db      427 cagaagcatatccctcattctctgtcgtatctgggggtgtgttctctgtgggtgagaagatgtc 446
```

```
Oy      4413 gtgccttgatacctgtgggagccctccacactcctctgaacccat 4455
```

```
Db      487 gttg-cgtgtcccaaccagatcagcttctctctgacccctt 528
```

## RESULT 15

```
AA562772
ID      AA562772 standard; cDNA; 722 BP.
```

```
AC      AA562772;
```

```
DT      14-FEB-2002 (first entry)
```

```
XX      cDNA sequence #559 encoding novel human secreted protein.
```

XX	Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW	immune deficiency disorder; blood disorder; inflammatory disorder;
KW	infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW	immunosuppressive; antirheumatic; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200177291-A2.
XX	
PD	18-OCT-2001.
XX	
PF	29-MAR-2001; 2001WO-US10485.
XX	
PR	06-APR-2000; 2000US-195604P.
XX	
PA	(GEMV ) GENETICS INST INC.
XX	
PI	Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI	Gaukkota K, Graham JR;
XX	
DR	WPI: 2002-010900/01.
XX	
PT	New polynucleotides encoding secreted proteins useful for treating e.g.
PT	asthma, HIV and Crohn's disease -
XX	
XX	
PS	Claim 1: Page 358; 391pp; English.
XX	
CC	The present invention relates to the isolation of novel cDNA sequences
CC	which encode human secreted proteins. The cDNA sequences have been
CC	derived from a variety of human tissues. The invention also provides
CC	a method for producing proteins from these polynucleotide sequences.
CC	The proteins are useful for identifying compounds that modulate their
CC	activity and production, and the cell is also useful for identifying
CC	compounds that modulate expression of the polynucleotide sequences
CC	encoding the secreted proteins. The sequences of the invention are
CC	useful for treating diseases such as hyperproliferative disorders
CC	(e.g. cancer), immune deficiency disorders (e.g. severe combined
CC	immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC	sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC	disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC	The polynucleotide sequences of the invention are also useful in gene
CC	therapy. AAS62214-AAS62838 represent the cDNA sequences of the
CC	invention that encode for novel human secreted proteins.
XX	
XX	
XX	

Query Match	2.3%	Score 109.4	DB 24	Length 722
Best Local Similarity	60.9%	Pred. No. 2.4e-20		
Matches 231	Conservative 0	Mismatches 141	Indels 7	Gaps 3
OY 4080	tcgagaagaggaatgagctgtgctcttlttgcaagcagcagcccaagaatactgcaaac	4139		
DB 174				
OY 4140	cctctgtatgagacctcagcgcgcgcacccacagagtcctctctgagcccttaacag	4199		
DB 234				
OY 4200	cctctgtatgagacctcagcgcgcgcacccacagagtcctctctgagcccttaacag	4256		
DB 294				
OY 4257	gagaagccaagcaagaaccacagtcacggttcacagctgatactccagcccaagctgtctc	4316		
DB 354				
OY 4317	atgagtcagtggttttaagaagatgacaacatcattaaagccaagcaactacacctcttgc	4376		
DB 470				
OY 4377	gttaactctggagattctgcaccagctcctaaggaaccaattgtacctgataactatggagact	4436		
DB 470				

Db 471 gctgactcgggggtcgtgcttcctgtagagatgltgtcctggccacacagctacag 529

Search completed: June 17, 2002, 16:44:47  
Job time: 9872 sec

Tue Jun 18 09:06:50 2002

us-09-842-930a-1.rng

---

Page 24

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2002, 13:20:30 : Search time 88.46 Seconds  
(without alignments)  
13067.502 Million cell updates/sec

Title: US-09-842-930A-1  
Perfect score: 4706  
Sequence: 1 tcttaccacagctctactcac.....aagcaaaaaaaaaaaaaaa 4706

Scoring table:  
IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/backfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.8	1.5	1414	1 US-08-024-868-1	Sequence 1, Appl
2	70.8	1.5	1414	2 US-08-242-097-1	Sequence 1, Appl
3	70.8	1.5	1414	4 US-09-206-695-1	Sequence 1, Appl
4	60	1.3	1400	2 US-08-001-078A-2	Sequence 2, Appl
5	60	1.3	1400	2 US-08-463-218-2	Sequence 2, Appl
6	60	1.3	1400	5 PCT-US94-00253-2	Sequence 2, Appl
7	44.2	0.9	7218	1 US-08-232-463-14	Sequence 14, Appl
8	43.4	0.9	7218	1 US-08-232-463-14	Sequence 14, Appl
9	41.2	0.9	1519	1 US-08-225-477B-2	Sequence 2, Appl
10	41.2	0.9	1519	5 PCT-US95-04353-2	Sequence 2, Appl
11	40.8	0.9	5191	1 US-08-340-428B-1	Sequence 1, Appl
12	40.8	0.9	5191	5 PCT-US93-07306-1	Sequence 1, Appl
13	40.4	0.9	8224	6 5180808-1	Patent No. 5180808
14	38.6	0.8	2461	1 US-08-282-141-1	Sequence 1, Appl
15	38	0.8	2049	1 US-08-268-797-1	Sequence 1, Appl
16	38	0.8	2049	5 PCT-US95-08414-1	Sequence 1, Appl
17	38	0.8	2691	1 US-07-878-960-1	Sequence 1, Appl
18	38	0.8	2892	1 US-08-264-534-5	Sequence 5, Appl
19	38	0.8	2892	1 US-08-083-590A-1	Sequence 1, Appl
20	38	0.8	2892	1 US-08-465-500-5	Sequence 5, Appl
21	38	0.8	2892	2 US-08-346-126-5	Sequence 5, Appl
22	38	0.8	2892	2 US-08-346-126-5	Sequence 5, Appl
23	38	0.8	2892	3 US-08-532-384-1	Sequence 1, Appl
24	38	0.8	2892	3 US-08-893-828-5	Sequence 5, Appl
25	38	0.8	3259	5 PCT-US95-03747-1	Sequence 1, Appl
26	37.4	0.8	3215	1 US-08-426-627-1	Sequence 1, Appl
27	37	0.8	1520	1 US-08-225-477B-1	Sequence 1, Appl

28	37	0.8	1520	5 PCT-US95-04353-1	Sequence 1, Appl
29	36.4	0.8	3250	1 US-07-906-395-1	Sequence 1, Appl
30	36.4	0.8	3250	1 US-08-192-632-1	Sequence 1, Appl
31	36.4	0.8	3250	1 US-08-710-676-1	Sequence 1, Appl
32	36.4	0.8	3250	3 US-09-099-902B-1	Sequence 1, Appl
33	36.4	0.8	3250	5 PCT-US93-06080-1	Sequence 1, Appl
34	36.2	0.8	289	4 US-09-007-005-17	Sequence 17, Appl
35	36.2	0.8	289	4 US-09-244-796-17	Sequence 17, Appl
36	36.2	0.8	18609	4 US-08-943-731-1	Sequence 1, Appl
37	36	0.8	975	6 5340934-10	Patent No. 5340934
38	34.6	0.7	2447	2 US-09-014-969-14	Sequence 14, Appl
39	34.4	0.7	977	6 5340934-9	Patent No. 5340934
40	34.4	0.7	3709	4 US-09-541-782-7	Sequence 7, Appl
41	34.4	0.7	6049	5 PCT-US95-11684-3	Sequence 3, Appl
42	33.8	0.7	4411529	4 US-09-103-840A-1	Sequence 1, Appl
43	33.6	0.7	804	2 US-08-611-880-4	Sequence 4, Appl
44	33.2	0.7	198	1 US-08-330-108-16	Sequence 16, Appl
45	33.2	0.7	198	5 PCT-US92-10087-16	Sequence 16, Appl

## ALIGNMENTS

RESULT 1  
US-08-024-868-1  
Sequence 1, Application US/08024868  
Patent No. 5386013  
GENERAL INFORMATION:  
APPLICANT: Lee, Tae Ho  
APPLICANT: Wisniewski, Hans Georg  
TITLE OF INVENTION: Cyclokinine-Induced Protein, TSG-6, DNA  
TITLE OF INVENTION: Coding Therefor and Uses Thereof  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Broadway and Neimark  
STREET: 419 Seventh Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: US  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/024,868  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,312  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Livanat, Shmuel  
REGISTRATION NUMBER: 33,949  
REFERENCE/DOCKET NUMBER: VILCEK-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-628-5197  
TELEFAX: 212-737-3528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1414 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: Fibroblast  
CELL LINE: FS-4  
FEATURE:  
NAME/KEY: CDS

LOCATION: 69, 899  
OTHER INFORMATION:  
US-08-024-868-1

Query Match 1.5%; Score 70.8; DB 1; Length 1414;  
Best Local Similarity 50.6%; Pred. No. 2.5e-10;  
Matches 171; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 3128 acggaagctgcacccagatgcagctgtgcagaccttaacttcagagagacagccgtag 3187  
DB 115 AAGGATGGGGATTCAAGATGAGATTTTCATTAACATCCATATGGCTTGAAGCAGCAGCCG 174  
QY 3188 gagatccatccatcagctcccccagctgcagagacaaactgacatttgacaagccaaag 3247  
DB 175 GTGTGTACCAAGAGAGACAGCTGTGGCAAAATCAAGCTACCTACGCGAAGAGCTTAAG 234  
QY 3248 aagcctgtgcacaaagactgtgcacatagccacactacacagcctcctctatgcacaga 3307  
DB 235 CGGTGTGATTTGAAGGCGGCCATCTCGCAACTTACAAGCAGTAAAGCAGCAGCA 294  
QY 3308 aggcacaagatagcctgtcctgcgcgcgtgtgagagatgggggttgctaccga 3367  
DB 295 AATTGGATTGATGTCTGTCTGTGATGATGGATGAGGACAGATTGGATACCCCA 354  
QY 3368 ctacgatatccctcagaagctgtgtgcacaaagctgtgtgagctgagactagacagatcca 3427  
DB 355 TTGTGAAGCAGGCGCCCACTGTGATTTGAAAAAAGTGGCATTTATGATGATTC 414  
QY 3428 gggccacaagaatgaatgtggatgtctctgttac 3465  
DB 415 GTCTCAATAGAGTGAAGATGGATGCTATTGTCTAC 452

RESULT 2

US-08-242-097-1  
Sequence 1, Application US/08242097  
Patent No. 5846763

GENERAL INFORMATION:  
APPLICANT: Lee, Tae Ho  
APPLICANT: Wisniewski, Hans Georg  
APPLICANT: Vilcek, Jan  
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding  
TITLE OF INVENTION: Therefor and Uses Thereof  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,097  
FILING DATE: 18-MAY-1994

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25, 618  
REFERENCE/DOCKET NUMBER: LEE26/VILCEK-1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-628-5197  
TELEFAX: 212-737-3528  
INFORMATION FOR SEQ. ID NO. 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1414 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: Fibroblast  
CELL LINE: FS-4  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 69, 899  
OTHER INFORMATION:  
US-08-242-097-1

Query Match 1.5%; Score 70.8; DB 2; Length 1414;  
Best Local Similarity 50.6%; Pred. No. 2.5e-10;  
Matches 171; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 3128 acggaagctgcacccagatgcagctgtgcagaccttaacttcagagagacagccgtag 3187  
DB 115 AAGGATGGGGATTCAAGATGAGATTTTCATTAACATCCATATGGCTTGAAGCAGCAGCCG 174  
QY 3188 gagatccatccatcagctcccccagctgcagagacaaactgacatttgacaagccaaag 3247  
DB 175 GTGTGTACCAAGAGAGACAGCTGTGGCAAAATCAAGCTACCTACGCGAAGAGCTTAAG 234  
QY 3248 aagcctgtgcacaaagactgtgcacatagccacactacacagcctcctctatgcacaga 3307  
DB 235 CGGTGTGATTTGAAGGCGGCCATCTCGCAACTTACAAGCAGTAAAGCAGCAGCA 294  
QY 3308 aggcacaagatagcctgtcctgcgcgcgtgtgagagatgggggttgctaccga 3367  
DB 295 AATTGGATTGATGTCTGTCTGTGATGATGGATGAGGACAGATTGGATACCCCA 354  
QY 3368 ctacgatatccctcagaagctgtgtgcacaaagctgtgtgagctgagactagacagatcca 3427  
DB 355 TTGTGAAGCAGGCGCCCACTGTGATTTGAAAAAAGTGGCATTTATGATGATTC 414  
QY 3428 gggccacaagaatgaatgtggatgtctctgttac 3465  
DB 415 GTCTCAATAGAGTGAAGATGGATGCTATTGTCTAC 452

RESULT 3

US-09-206-695-1  
Sequence 1, Application US/09206695  
Patent No. 6210905

GENERAL INFORMATION:  
APPLICANT: Lee, Tae Ho  
APPLICANT: Wisniewski, Hans Georg  
APPLICANT: Vilcek, Jan  
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA Coding  
TITLE OF INVENTION: Therefor and Uses Thereof  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/206,695  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,097  
FILING DATE: 13-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25, 618



```

? REFERENCE/DOCKET NUMBER: LEE26/YILICER-1B
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-628-5197
? TELEFAX: 212-737-3528
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1414 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? ORIGINAL SOURCE:
? ORGANISM: Homo sapiens
? CELL TYPE: Fibroblast
? CELL LINE: FS-4
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 69..899
? OTHER INFORMATION:
? US-09-206.695-1

```

[illegible]

```

1      RESULT      4
2      US-08-001-078A-2
3      : Sequence 2, Application US/08001078A
4      : Patent No. 5872094
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Goettnck, Paul F.
9      : APPLICANT: Tondravi, M., Mehرداد
10     : APPLICANT: Binette, Francois
11     :
12     : TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX
13     :
14     : TITLE OF INVENTION: FORMATION
15     :
16     : NUMBER OF SEQUENCES: 2
17     :
18     : CORRESPONDENCE ADDRESS:
19     :
20     : ADDRESSEE: LAHIVE & COCKFIELD
21     :
22     : STREET: 60 STATE STREET, Suite 510
23     :
24     : CITY: BOSTON
25     :
26     : STATE: Massachusetts
27     :
28     : COUNTRY: USA
29     :
30     : ZIP: 02109
31     :
32     : COMPUTER READABLE FORM:
33     :
34     : MEDIUM TYPE: Floppy disk
35     :
36     : COMPUTER: IBM PC compatible
37     :
38     : OPERATING SYSTEM: PC-DOS/MS-DOS
39     :
40     :

```

```

SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/001,078A
FILING DATE: 06-JAN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MPD-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

	Query Match	1.3%;	Score 60;	DB 2.0;	Length 1400;
	Best Local Similarity	52.7%;	Pred. No. 3.3e-07;		
	Matches 157;	Conservative	0;	Mismatches 135;	Indels 6;
					Gaps
QY	3172	caggacacgacgylaggagatattccatctacgctcccccacgtggccagctacaaactgaca	3233		
Db	694	CTGACCTTACAAAGGTGTGGCTATTCCCTTACTTTCACAGACTGGGGCGCTACAAATCTCAAT	753		
QY	3232	tttacacaagcacaagaagagcctgtgcacaagaagctcgtgacccatagcacctacaaccag	3291		
Db	754	TTTACACAGGCGCAGCAGGGCTCTGTGGACCAAGATGTGTATGCCCTCTTTCGACGAG	813		
QY	3292	ctctctcatgtgccagaagagcccaaggtatcacctgtgtcgtcgcgctgtgtctgtgaagtggg	3351		
Db	814	CTGTACAGACGCTGGCGGGCGGGGCGCTGGACTGTGGCAATGCCCGCTGGCTCATGATATGCC	873		
QY	3352	cgggtttgcctaacccgactatagatgtcctctccacaagaatgtggtgtg-----caaaqatgtgt	3405		
Db	874	TCTGTGCAATCTCCCATTCACAAAGCCACAGAGCCCTGTGGGGGCCCAAGAACACAGTGGCC	933		
QY	3406	ggagatcgtagactacaggaatccaggtgccacaagaagtgaagaatgtgggaatgtctcttcgtt	3463		
Db	934	GGAGTCCAGGAATCTACGGATTTTGTGGATTAAGAATAAAGACAGATATGATATTTTCTGTT	991		

RESULT 5  
 US-08-463-218-2  
 : Sequence 2, Application US/08463218  
 : Patent No. 5986052  
 : GENERAL INFORMATION:  
 : APPLICANT: Goettnck, Paul F.  
 : APPLICANT: Gonttravi, Mehrida  
 : APPLICANT: Binette, Francois  
 : TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX  
 : FORMATION  
 : NUMBER OF SEQUENCES: 2  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: LAHIVE & COCKFIELD  
 : STREET: 60 STATE STREET, Suite 510  
 : CITY: BOSTON  
 : STATE: Massachusetts  
 : COUNTRY: USA  
 : ZIP: 02109  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: ASCII text  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/463, 218  
 : FILING DATE: 05-JUNE-1995  
 : PRIOR APPLICATION DATA:  
 :

Query Match	1.3%	Score 60;	DB 5;	Length 1400;
Best Local Similarity	52.7%;	Pred. No. 3.3e-07;		
Matches 157; Conservative	0;	Mismatches 135;	Indels 6;	Gaps 14

Query Match	Score	DB	Length
0.98;	44.2;	1;	7218;

```
Best Local Similarity 11.1%; Pred. No. 0.032;
Matches 52; Conservative 202; Mismatches 215; Indels 0; Gaps 0;

QY 857 ggcgaggaggaatgtagatgagatcacacagacatgcaacggcgcgtgca 916
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1413 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1354

QY 917 ccaatgcgaactgctcttgatccagcagcgaagcctctgcaaatgctgcgagat 976
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1353 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1294

QY 977 tcgaggaatgtaacggtctgcacagcatcaatgctgtgaacacgaatgagat 1036
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1293 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1234

QY 1037 gtctacaagcgcgactgtaaaagaaccaccaggaacgggtgtgtgtgcagag 1096
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1233 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1174

QY 1097 caggctataccgcgcagcgcacgtgtgtctgaatcaaccgtgtgtgagaacatg 1156
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1173 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1114

QY 1157 gtggtcttgacaagaatgcaagtgacacagacagagcccaacgagcgctgtact 1216
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1113 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1054

QY 1217 gcttcgcaaatacacgaggaatgaaagtgctgtcgtctatcaatgtctgcctaa 1276
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1053 TCGACCTGCACCAAGCTCGCAATTATTCTGTGAGCGTATGGCAACGAAGAAATA 994

QY 1277 acaatgagcgtgcatcattgtcctctctgcaactacactgacgaaga 1325
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 GTTATTAGACCGCAGCTCGATGGACATTTCACGTAACGTTAATA 945

RESULT 8
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232.463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935.313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
```

```
TELEFAX: (703)683-4109
:
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-F15
US-08-232-463-14

Query Match 0.9%; Score 43.4; DB 1; Length 7218;
Best Local Similarity 3.6%; Pred. No. 0.054;
Matches 14; Conservative 212; Mismatches 163; Indels 0; Gaps 0;

QY 4259 gaagcagaagaacacacagctcacggtcacaggtgtccagccagcgtctcat 4318
    ||| ||| : : : : : : : : : : : : : : : : : : : :
Db 1055 GAGGCTTCGATMYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1114

QY 4319 ggaatcagttttaaagaatgacacactcataagccagcataccctctcgtg 4378
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1115 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1174

QY 4379 taatcggatgttcgcagcagcctaagagccatgttcctgatacctgggagctcc 4438
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1175 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1234

QY 4439 acctcctctgacgtacacgtgtctctcactcactcactatgctgtctgtctgc 4498
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1235 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1294

QY 4499 cctctctgtaccccaacacgtgactctgtggtatctctcatgacgtgaagcccaa 4558
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1295 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1354

QY 4559 ggcgggctcacctcctatgtctglatctcagtaaccagaagtaacctgcacacatg 4618
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1355 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1414

QY 4619 gtgctcaataatgtttgtggaacaaat 4647
    : : : : : : : : : : |||||
Db 1415 YYYYYYYYYYYYYYYYYYGTACCAAAAT 1443

RESULT 9
US-08-225-477B-2
: Sequence 2, Application US/08225477B
: Patent No. 5635370
: GENERAL INFORMATION:
: APPLICANT: Susan Hockfield
: APPLICANT: Diane M. Jaworski
: TITLE OF INVENTION: BEHAV. A Brain Hya-
: TITLE OF INVENTION: Juruman-Binding Protein
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: St. Onge Steward Johnston & Reens
: STREET: 986 Bedford Street
: CITY: Stamford
: STATE: CT
: COUNTRY: United States
: ZIP: 06905
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" 1.44 Mb diskette
: COMPUTER: IBM PC
: OPERATING SYSTEM: MS DOS
: SOFTWARE: Word Processor
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/225.477B
: FILING DATE: April 8, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Mary M. Krinsky
```



```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5191 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 77..3847
; US-08-340-428b-1

Query Match          0.9%: Score 40.8; DB 1; Length 5191:
Best Local Similarity 49.5%; Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 3160 gacctctactccagagacagcgttagagatltccatcctcccactggccag 3219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 GACCTGTGTAACCCGTGGAGAGTACGGCGCTGTTCCATTATCGGGCGCCCGGACCGC 583

QY 3320 tacaactgacatttgacaagaagcctgtgccaagaagctgcgaccataagc 3279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 TATCGTTGACCTTCGGTGGAGGCCAGAGGCTGTCTACCTGACCTCCGCTACCATTTGCG 643

QY 3280 acctacaacagcgtctctctatgcccagaagccaagtatcacctgtgctgcgcgcttg 3339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 644 GCTCCAGAGCGACCTGCGAGCGCTGCTTCGACATGATGCTTGACAACTCGCATGCGGGCTGG 703

QY 3340 ctggagagtggtggcggtgtgctaccgactac 3371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 704 CTCTCAGACCGCAGCGTCCGGTACCCGATCAC 735

RESULT 12
PCT-US93-07306-1
; Sequence 1, Application PC/TUS9307306
; GENERAL INFORMATION:
; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Uwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07306
; FILING DATE: 03-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/922,911
; FILING DATE: 03-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Guy K.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: Margolis-1A PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5191 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 77..3847
; PCT-US93-07306-1

Query Match          0.9%: Score 40.8; DB 5; Length 5191:
Best Local Similarity 49.5%; Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 3160 gacctctactccagagacagcgttagagatltccatcctcccactggccag 3219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 GACCTGTGTAACCCGTGGAGAGTACGGCGCTGTTCCATTATCGGGCGCCCGGACCGC 583

QY 3320 tacaactgacatttgacaagaagcctgtgccaagaagctgcgaccataagc 3279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 TATCGTTGACCTTCGGTGGAGGCCAGAGGCTGTCTACCTGAGCTCCGCTACCATTTGCG 643

QY 3280 acctacaacagcgtctctctatgcccagaagccaagtatcacctgtgctgcgcgcttg 3339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 644 GCTCCAGAGCGACCTGCGAGCGCTGCTTCGACATGATGCTTGACAACTCGCATGCGGGCTGG 703

QY 3340 ctggagagtggtggcggtgtgctaccgactac 3371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 704 CTCTCAGACCGCAGCGTCCGGTACCCGATCAC 735

RESULT 13
5180808-1
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO:1
; LENGTH: 8224
; 5180808-1

Query Match          0.9%: Score 40.4; DB 6; Length 8224:
Best Local Similarity 48.7%; Pred. No. 0.43;
Matches 110; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 3190 gtattcattcactgcctcccaactggccagatcacaaacttgagaaagccaagaag 3249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 717 gtgttccactacagggggcgaacagcaggtacacacgaatttggagctgcgtcagaag 776

QY 3250 gctgtgccaagaagctgcgacataagcaccctacaaacagctctcctatgccaagaag 3309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 777 gctgttgtagcgttggtggcagctatagcaactccagagcagctcttgcgtcctatgaa 836

QY 3310 gccaaatcacctgtgctgcgcgctgctgtgagaagtggggcgtgtgctctaccagact 3369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 837 gatgatttgacagctgtgacagcagctgcgtgtgatacagactgtcatagatatccatc 896

QY 3370 acgtatgctctccagaagtgtgctgcaaaagttgttggatcgtag 3415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 897 cgggtctcccaagatgactgttattgagatagaatgggaaagcgag 942

RESULT 14
US-08-282-141-1
; Sequence 1, Application US/08282141
```











FT	Domain	943..957	/note= "EGF-like type 2 domain"
FT	Domain	943..977	/note= "laminin-type EGF domain"
FT	Domain	987..998	/note= "EGF-like type 2 domain"
FT	Domain	987..998	/note= "EGF-like type 2 domain"
FT	Domain	1027..1040	/note= "EGF-like type 1 domain"
FT	Domain	1069..1082	/note= "EGF-like type 2 domain"
FT	Domain	1103..1113	/note= "EGF-like type 2 domain"
FT	Domain	1111..1125	/note= "prokaryotic membrane lipoprotein lipid attachment site domain"
FT	Domain	1405..1415	/note= "EGF-like type 2 domain"
FT	Domain	1582..1616	/note= "prokaryotic membrane lipoprotein lipid attachment site domain"
FT	Domain	1582..1596	/note= "EGF-like type 2 domain"
FT	Domain	1582..1593	/note= "EGF-like type 1 domain"
FT	Domain	1626..1637	/note= "EGF-like type 2 domain"
FT	Domain	1626..1637	/note= "EGF-like type 1 domain"
FT	Domain	1663..1676	/note= "EGF-like type 2 domain"
FT	Domain	1747..1760	/note= "EGF-like type 2 domain"
FT	Domain	1791..1894	/note= "HA binding motif"
FT	Domain	1817..1862	/note= "link protein domain"
FT	Domain	1894..1908	/note= "EGF-like type 2 domain"
FT	Misc-difference	2058	/note= "unspecified amino acid encoded by Cnr"
FT	Misc-difference	2109	/note= "unspecified amino acid encoded by GAN"
FT	Misc-difference	2123	/note= "unspecified amino acid encoded by Tnr"
XX	WO200039166-A1.		
PD	06-JUL-2000.		
PF	20-DEC-1999;	99WO-US30462.	
PR	23-DEC-1998;	98US-0113871.	
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(AMNA-) AMERICAN NAT RED CROSS.		
PI	Hastings GA, Llau G, Tsiftina E;		
DR	WP: 2000-452376/39.		
DR	N-PSDB: AAA57362.		
XX			
PT	OE-hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP		
PT	metastasis, inflammation, ischemia, arthritis and multiple sclerosis -		
XX	Claim 11; Fig 1-H; 457pb; English.		
XX			
CC	The present sequence represents a hyaluronan-binding protein. The		
CC	specification describes four hyaluronan-binding protein, known as		
CC	WF-HABP, WF-HABP, OE-HABP, and BW-HABP. The polypeptides are useful		
CC	for treating diseases such as proliferative conditions, metastasis,		

[illegible]

```
Dh 1600 afghbqacrtctvhyrdecldgsgscfcdgwtgtrcevgldlqpvctppcpeavcra 1659
Qy 929 NNTVCVCLANESGIGITCTVVDFFCKQNNGCACAKVAKSOKGTQVSCSKKTKGKGCYCIE 988
Dh 1600 gnscecslyegvdgrvtvadlqgdhngscsehnscsqvgltmvtctclpdyegdwscra 1719
Qy 989 IDPCADGVNCGCHEHATCRMTGPGKHKCEKSHYVGDVDC-EPEQCLPDRCLQDNQCCH 1047
Dh 1720 rncptdghrgscsehnclstglnttrcechagvygdqlcleesepvdcltqgppch 1779
Qy 1048 PDASCADLYTQDTTVYFHLRSPGLQYKLFDPKAKECAKEAATIAIYVNDLSYAKAKYH 1107
Dh 1780 sdcmclthfgekragvfhlgatsgpyglhfseaeaceagavlasfpqisaagqlgfh 1839
Qy 1108 LGSAGMESGRVAPPTTYASOKGCANVVGIVDYSRANKSPMVMFCYRKMDVCTCKAG 1167
Dh 1840 lclmgwlaangslanpvhfvpadcngtrgvsvlgarkhlsrwayclrvdvcrcrnq 1899
Qy 1168 YVSDGFES-CSGNLLOVLMSPFSLTNFLTEVLAFESKSARGQAFLEKHLTDLISRGTFEPQ 1226
Dh 1900 fvgdglstcngklldvlaacantfcfygmllgyanacqrgldfidldeltykltltpv 1959
Qy 1227 NSGLPGKSLSGRDIIEHLTNVNSFYNDLVNGTFLRTMLGSQLLTFPSQDLHQETRFV 1286
Dh 1960 negfvdmtlspnllehasnatllsan-asgqklpahsgllslsdagpdmsswapva 2018
Qy 1287 DG-----KSIIOWDILANGILHITSEPLRAP-----TAATAASGLGTGIFCAVLVTG 1337
Dh 2019 pgtvsvrsltlwldmalngllhalaspdlapppqavlaaxepvaagvavlaaga1lg 2078
Qy 1338 AIALAAYSFRLKORTGPORF 1359
Dh 2079 lvaqalyllarqkpmgfgsaf 2100

RESULT 2
AAB83364
ID AAB83364 standard; Protein; 669 AA.
XX
AC AAB83364;
XX
DT 26-MAR-2002 (first entry)
DE
DE NOV8 protein sequence.
XX
KW NOV: Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
KW cardiovascular; casein kinase II phosphorylation site; contraction;
KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
KW epidermal growth factor; cell development; apoptosis; cell adhesion;
KW growth migration; cell structure; motility; cancer; immune disorder;
KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.
XX
OS Unidentified.
XX
PN MO200136638-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000MO-US31543.
XX
PR 19-NOV-1999; 99US-166336P.
PR 29-NOV-1999; 99US-167785P.
PR 08-MAR-2000; 2000US-187844P.
PR 16-NOV-2000; 2000US-0715417.
XX
XX (CURA-) CURAGEN CORP.
PA
XX Shimkets RA, Lichenstein H, Vernet C, Fernandes E;
PI
DR MPI; 2001-648134/74.
```

```
DR N-PSDB; AAF87119.
XX
PT Novel human polypeptides and the nucleic acids that encode them useful
PT for preventing, diagnosing and treating e.g. cancer, inflammation and
PT immune disorders -
XX
PS Claim 1; Page 29-30; 141pp; English.
XX
CC This sequence is the NOV8 protein. The invention relates to
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC cytosolic; contraceptive; antiinflammatory; immunomodulatory; and
CC cardiovascular activities. The sequences may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NOVX
CC expression. They may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of protein by expressing inactive proteins or to
CC supplement the patients own production of protein. They are used to
CC produce NOVX proteins, by inserting the nucleic acid into a cell and
CC culturing it to express the protein. The DNA may be used as DNA probes in
CC assays to detect and quantitate the presence of similar DNAs in samples,
CC and which patients may need restorative therapy. The NOVX protein may
CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonist are used to down regulate
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contraction). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovascular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX
SQ Sequence 669 AA;

Query Match 34.5%; Score 2710.5; DB 22; Length 669;
Best Local Similarity 78.8%; Pred. No. 1,5e-175;
Matches 471; Conservative 60; Mismatches 66; Indels 1; Gaps 1;
```

QY 929 NMTCVONLWVEGGSTCTTCAVDPECKNNNGGAKATKACQKQTQVSCSCKKCYKGNGYCIE 988

Db 487 mtteccnlalveyggctictvddictkqdnngagcavarsqgklvkscsqgkykgdghacte 546

QY 989 IDPCADGVNAGGGEIHAHTCWMTGCGKHKECEKSHYVGCVGVCDEEQLDLPBCLJONGCCHP 1048

Db 547 idpcadvglingelhehtckmtktmtpgpkhceekshyvgdglncepeqlpidrcldngnqga 606

QY 1049 DASCADLYRQDFIVGYFHLRLSPGLGYKLFLFDKAKEACAEATATATYNQLSYAOKARY 1106

Db 607 dakcvdlhtgdetevgyfhlrpslpgykllftdkareacanaatmatynqlsyagtkwy 664

RESULT 3

AAB83362

ID AAB83362 standard Protein; 669 AA.

XX AAB83362;

XX DT 26-MAR-2002 (first entry)

XX NOV6 protein sequence.

XX NOV, Cytostatic; contraceptive; antiinflammatory; immunomodulatory;

KM cardiovascular; casein kinase II phosphorylation site; contraception;

KM serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;

KM epidermal growth factor; cell development; apoptosis; cell adhesion;

KM growth migration; cell structure; motility; cancer; immune disorder;

KM inflammatory disorder; cellular adhesion disorder; long-QT syndrome;

KM cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;

KM therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;

XX NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.

OS Unidentified.

PN WO200136638-A2.

PD 25-MAY-2001.

XX 17-NOV-2000; 2000WO-US31543.

PF 19-NOV-1999; 99US-166336P.

XX 29-NOV-1999; 99US-167785P.

PR 08-MAR-2000; 2000US-187844P.

PR 16-NOV-2000; 2000US-0715417.

XX (CURA-) CURAGEN CORP.

PA Shimkets RA, Lichenstein H, Vernet C, Fernandes E;

PI WPI: 2001-648134/74.

DR N-PSDB; AAF87117.

XX Novel human polypeptides and the nucleic acids that encode them useful

PT for preventing, diagnosing and treating e.g. cancer, inflammation and

PR immune disorders.

XX Claim 1; page 24-25; 141pp; English.

PS This sequence is the NOV6 protein. The invention relates to

CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have

CC cytostatic; contraceptive; antiinflammatory; immunomodulatory; and

CC cardiovascular activities. The sequences may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate NOVX

CC expression. They may be used to treat disorders associated with decreased

CC that affect the activity of protein by expressing inactive proteins or

CC supplement the patients own production of protein. They are used to

CC produce NOVX proteins, by inserting the nucleic acid into a cell and

CC culturing it to express the protein. The DNA may be used as DNA probes in

CC assays to detect and quantitate the presence of similar DNAs in samples,

CC and which patients may need restorative therapy. The NOVX protein may

Query Match	34.4%	Score 2701.5	DB 22	Length 669
Best Local Similarity	78.6%	Pred. No. 6.3e-175		
Matches 470	Conservative 60	Mismatches 67	Indels 1	Gaps 1
509 RIKMDQGLMSQVLRHYHVGCOQLLDMKVTTSATTLGGSPVSISSQDTVFINNNAK	568			
68 gkdwkkygimpqvltfhyvachqlllenklklnatslqgprlvasqvetylnnak	127			
569 VLSSDITISNGVIVHIDKLISPKNLLITPKDALGVLOWLTVVAANHGTKRSKSLQDSG	628			
128 lsssdlsfngvthidkllspknllltpkdnsgylllqltllatnngyikfslndqsg	187			
629 LLSVTIDSHTFTVTFWFPDKALEALPEPQOFLFNODKDKLSYLNKHNVRIRDSKALAS	688			
188 llsvltqphtrpytlfwpdgalhaipaegqfllfngdkkdklkeylkhvltirdakvlay	247			
689 DLPKRSAMKTLGGSELSESVRCSTGSGSDIGELFLNEQMCRTFNRGLRFDVGYAVIDCLLMP	748			
248 dlptstawktlqgselsvkcqgrdlgdlfngqlcrliqvrelldlgyayidcllldp	307			
749 TLGGRCDFTTTDDITGEGSCSCTFTFKRCPLSKSPKGVKKKCIYNPLPRFNNVGGQMLCIV	808			
308 tlfgtrcdtlltldasqegscvnlrpscpwskpkgykqclyn-lpikfnlegcrecs1	366			
809 VIQTPRCCHGYFMPDQOACPGSPDTPCNNRGMCRDLYTPMGQCLCHTGTGNFACELCHNG	868			
367 vqiprcckgytgrdqacp9p9vapecmrtycl0dyselgeckrclyfnglacemcpw9	426			
869 RFGPDQPRSCSEHGQCDGEGINGSEBCLCETGWTAA5CDTPPAVFAVCPACSVHATTE	928			
427 rtfqpdclpogcsghgqcdgdlqsgsqclcetqwtvtpscdtqglv savctprpsahatcke	486			
929 NNTVCNLTNEEDDGIITYVDFCRKONNGCACAIVAKCSQKTOVSSCSCKKGYGDSCTE	988			
487 mlteencndlyegdqtlctvdfcdkdnngcavarcosqgkltvscsqkgygqgdhsche	546			
989 IDPCADGVNGGCHENATCRMTGPKGNKRCCKSHYVGDGVDCEPPEQDLPDRCLQDNQGCIP	1048			
547 idpccadglnghenahatckmtgpkghkceckshyvgdqlncepqldlrcldqanqgcha	606			
1049 DASCADLFFQDTTVGVFHLRSPILGQYKLTDFDKAKACAKAEATLTVNQLSYAORAKY	1106			
607 dekcvdlhbgdtctvgvfhrlspilgqykltdkaraecacneaatmtctnqlsyagqktwy	664			

[illegible]

D6  
61 rsplgylkltfdkareacaneaetmatltnqslsyxqakayhlcsagmyelgrvaprtafas 120

OY 1128 QKCGANWVIVDYGSFRANKSEMDVFQCYRMKDYNCTCKAGAYGDGFGSCGNLLQVLMSFP 1187  
+ + : ||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||  
D6 121 qncsgvgivdygprrpnksmwdvfecymkdvcnckvkgyvdgtfsysgnllqvlnmsfp 180

OY 1188 SLTNNFLTEVLAAFSKSARQAELFKHLTDLSIKRTLFVPONSGLPGNKSLSGRDIENHILT N 1247  
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
D6 181 slntflteyLaysnsasargaflehtldislrgtlftvpqnsqigenetclisrghelan 240

OY 1248 VWSFNPNDLVNGCFLEPTMTLGSQLTFEFDSDOLH-OETRFVDGSTLIOWPDI IANGJILHIIT 1306  
+ + + + + |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||  
D6 241 vmfflyndlvngtclqlctrfgsklllldrqdpprlpletrecvdgdctlewdlcasngilhv l 300

OY 1307 SEPLRAPPPAATAAHSGLTGTGFCAVALVTGAIALAAYSFYRLKORTTGQORF 1359  
+ + : ||| | | : ||| | | : ||||| : ||||| : ||| | |  
Db 301 srxlkappavvlxbhgjlgxfmxllvtgavalaaesyfrlnrkltigfzhf 353

RESULT 5  
AAB83358 standard; Protein; 315 AA.  
ID AAB83358 standard; Protein; 315 AA.

XX AAB83358;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DE NOV2 protein sequence.

KM NOV. Cytostatic; contraceptive; antiinflammatory; immunomodulatory;  
KM cardiovascular; casein kinase II phosphorylation site; contraception;  
KM serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;  
KM epidermal growth factor; cell development; apoptosis; cell adhesion;  
KM growth migration; cell structure; motility; cancer; immune disorder;  
KM inflammatory disorder; cellular adhesion disorder; long-QT syndrome;  
KM cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;  
KM therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;  
KM NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.

OS Unidentified.  
XX  
XX WO200136638-A2.  
PN  
PD 25-MAY-2001.  
XX  
PF 17-NOV-2000; 200OMO-US31543.  
XX  
PR 19-NOV-1999; 99US-166336P.  
PR 29-NOV-1999; 99US-167785P.  
PR 08-MAR-2000; 2000US-187844P.  
PR 16-NOV-2000; 2000US-0715417.  
XX  
PA (CURA-) CURAGEN CORP.

PX Shinkets RA, Lichenstein H, Vernet C, Fernandes E;  
PI  
XX WPI: 2001-648134/74.  
DR N-PADB; AAF87113.  
XX  
PT Novel human polypeptides and the nucleic acids that encode them useful  
for preventing, diagnosing and treating e.g. cancer, inflammation and  
immune disorders -  
PT  
PX  
PS Claim 1; Page 10-13; 141pp; English.

CC This sequence is the NOV2 protein. The invention relates to  
the NOV1-NOV16 proteins, and their coding sequences. The proteins have  
cytostatic; contraceptive; antiinflammatory; immunomodulatory; and  
cardiovascular activities. The sequences may be used in the prevention,  
diagnosis and treatment of diseases associated with inappropriate NOVX  
expression. They may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of protein by expressing inactive proteins or to

CC supplemented the patients own production of protein. They are used to  
CC produce NOVX proteins, by inserting the nucleic acid into a cell and  
CC culturing it to express the protein. The DNA may be used as DNA probes in  
CC assays to detect and quantitate the presence of similar DNAs in samples,  
CC and which patients may need restorative therapy. The NOVX protein may  
CC also be used as antigens in the production of antibodies (Abs) against  
CC NOVX and in assays to identify modulators of NOVX expression and  
CC activity. The anti-NOVX Abs and antagonist are used to down regulate  
CC expression and activity. The anti-NOVX Abs are used for detecting the  
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed  
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,  
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites  
CC characteristic of serine/threonine kinases, and are used to treat  
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular  
CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to  
CC the epidermal growth factor (EGF)-like super family and are involved in,  
CC e.g. regulation of cell development, apoptosis, cell adhesion, growth  
CC migration, cell structure and motility and protein management, and are  
CC used to treat cancers, inflammatory disorders, immune disorders and  
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like  
CC fibroillin proteins and are used to treat cardiovascular disease e.g.  
CC hypertrophic cardiomyopathy, Long-QT syndrome and marfan syndrome.  
XX

Query Match	18.98;	Score 1487;	DB 22;	Length 315;
-------------	--------	-------------	--------	-------------

Matches	256;	Conservative	24;	Mismatches	31;	Indels	16;	Gaps	1;
---------	------	--------------	-----	------------	-----	--------	-----	------	----

OY	864	LCGARGRGPDOGRSSCEHOCDEGTITGSECI.CELGMWAA.SODP.TPAFAFACPRAGSVH	923
Dd	1	mewgyr1gppdd1pgocsdhdngcoddgl1gsygc1.ccegmw1gppscdd1qav1asvcttprcsah	60
OY	924	ATCTENNTVCNLNLEYEGSDGITCTVDFCKONNGCAKVAKSQKTGYSCSKKKYGKDGG	983
Dd	61	atckentntccenldlyegdgdlctctvwdfckdqndgcaekvarcsqkylkvscsqkykgkdg	120
OY	984	YSCFEIDPCADGVNCGGHEHAHTCMTGPGKHKECKESHVVGSGVDPECPQLDLDICLDDN	1043
Dd	121	hsceidrcpadg1nnggchehatckmtgpyrkfhkeckshtvyvgd1lnepeqr1drclcdqn	180
OY	1044	GQCHPDASCADLYFDOTTVGCVPHLRSLFLGOYLKTFEDKAEACAKEAATATATYNOLSTVAOK	1103
Dd	181	gqhcadackadlhbgdtcttvgrfhlrsp1gryklktfokareacaneatalmatynq1stysaqk	240
OY	1104	AKYHLCSAGMLSESRRVAVPTTYASOKGAWVIVDYGSRAKNSKEMADVFRCRMADVCT	1163
Dd	241	akylhcsagvyletgrvaypafafasqmcsgsvyvgivdygprtksemwdvfcyrmymg----	296
OY	1164	CKAGYVGDGFS CSGNLLQVTLMSPSLT	1190
Dd	297	-----saglftqjssrfpcsls311	

RESULT	6
AAB83365	
ID	AAB83365 standard; Protein; 545 AA.

DT	26-MAR-2002	(first entry)
XX		
DE	NOV9	protein sequence.

KM NOV; Cystostatic; contraceptive; antiinflammatory; immunomodulatory;  
KM cardiovascular; casein kinase II phosphorylation site; contriception;  
KM serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation  
KM epidermal growth factor; cell development; apoptosis; cell adhesion;  
KM growth migration; cell structure; motility; cancer; immune disorder;  
KM inflammatory disorder; cellular adhesion disorder; long-QT syndrome;  
KM cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;  
KM therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;

KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.

OS Unidentified.

PN W0200136638-A2.

PD 25-MAY-2001.

17-NOV-2000; 2000WO-US31543.

PR 19-NOV-1999; 99US-166336P.

08-MAR-2000; 2000US-187844P.

XX

[illegible][illegible]

DR MF1, 2001 040134/11.  
DR N-PSDB; AAF87120.

AA	Novel human polyp
PT	

PT immune disorders

aa  
ps  
Claim 1: Page 32-33; 141pp; English.

This sentence is the NOV9 protein. The invention relates to the NOV1-NOV16 proteins, and their coding sequences. The proteins have Cytostatic; contraceptive; antiinflammatory; immunomodulatory; and cardiovascular activities. The sequences may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate NOVX expression. They may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of protein by expressing inactive proteins or to supplement the patient's own production of protein. They are used to produce NOVX proteins, by inserting the nucleic acid into a cell and culturing it to express the protein. The DNA may be used as DNA probes in assays to detect and quantitate the presence of similar DNAs in samples, and which patients may need restorative therapy. The NOVX protein may also be used as antigens in the production of antibodies (Abs) against NOVX and in assays to identify modulators of NOVX expression and activity. The anti-NOVX Abs and antagonist are used to down regulate expression and activity. The anti-NOVX Abs are used for detecting the presence of NOVX in samples. Disorders that may be prevented, diagnosed and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5, NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites characteristic of serine/threonine kinases, and are used to treat kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular proliferation and contraception). NOV-3, NOV6 and NOV9 are homologous to the epidermal growth factor (EGF)-like super family and are involved in, e.g. regulation of cell development, apoptosis, cell adhesion, growth migration, cell structure and motility and protein management, and are used to treat cancers, inflammatory disorders, immune disorders and cellular adhesion disorders. NOV6-10 are homologous to EGF-like fibillin proteins and are used to treat cardiovascular disease e.g. hypertrophic cardiomyopathy, Long-QT syndrome and marfan syndrome.

Query Match	16.88;	Score	1318.5;	DB	22;	Length	545;
Best Local Similarity	75.78;	Pred. No.	4e-81;				
Matches	240;	Conservative	37;	Mismatches	39;	Indels	1;
						Gaps	1

[illegible]



XX unidentified.  
OS  
XX WO200136638-A2.  
PN  
XX  
XX 25-MAY-2001.  
PD  
XX  
XX 17-NOV-2000; 2000MO-US31543.  
PF  
XX  
XX 19-NOV-1999; 99US-166336P.  
PR 29-NOV-1999; 99US-167785P.  
PR 08-MAR-2000; 2000US-187844P.  
PR 16-NOV-2000; 2000US-0715417.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX  
XX Shinkets RA, Lichenstein H, Vernet C, Fernandes E;  
PI  
XX WPI: 2001-648134/74.  
DR N-PSDB: AAF87114.  
XX  
XX Novel human polypeptides and the nucleic acids that encode them useful  
PT for preventing, diagnosing and treating e.g. cancer, inflammation and  
PT immune disorders -  
PT  
XX  
XX Claim 1: Page 14-17; 141pp: English.  
XX  
XX This sequence is the NOV3 protein. The invention relates to  
XX the NOV1-NOV6 proteins, and their coding sequences. The proteins have  
XX cytoskeletal; contractile; antiinflammatory; immunomodulatory; and  
XX cardiovascular activities. The sequences may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate NOVX  
XX expression. They may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of protein by expressing inactive proteins or to  
XX supplement the patient's own production of protein. They are used to  
XX produce NOVX proteins, by inserting the nucleic acid into a cell and  
XX culturing it to express the protein. The DNA may be used as DNA probes in  
XX assays to detect and quantitate the presence of similar DNAs in samples,  
XX and which patients may need restorative therapy. The NOVX protein may  
XX also be used as antigens in the production of antibodies (Abs) against  
XX NOVX and in assays to identify modulators of NOVX expression and  
XX activity. The anti-NOVX Abs and antagonist are used to down regulate  
XX expression and activity. The anti-NOVX Abs are used for detecting the  
XX presence of NOVX in samples. Disorders that may be prevented, diagnosed  
XX and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,  
XX NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites  
XX characteristic of serine/threonine kinases, and are used to treat  
XX kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular  
XX proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to  
XX the epidermal growth factor (EGF)-like super family and are involved in,  
XX e.g. regulation of cell development, apoptosis, cell adhesion, growth  
XX migration, cell structure and motility and protein management, and are  
XX used to treat cancers, inflammatory disorders, immune disorders and  
XX cellular adhesion disorders. NOV6-10 are homologous to EGF-like  
XX fibillin proteins and are used to treat cardiovascular disease e.g.  
XX hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.  
XX  
XX Sequence 244 AA:

Query Match 15.4%; Score 1209; DB 22; Length 244;  
Best Local Similarity 84.2%; Pred. No. 3.7e-74;  
Matches 202; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

QY 864 LCMHGFEGDQPRSCSEHGOCDEGTSGECICETGWTAAASODPTAAFAVCTPACSVH 923  
DB 1 mcvwprfipgddclpcgscsdngqcdgdgtlgsqgcetcgwqpsodtqavlsavctlpcpsah 60  
QY 924 ATCTENNITVCVNLINEGSDGITCTVVDGFCQONNGCAKVAKCSQKGTQVSCGCKKGYKGG 983  
DB 61 atckenntecndlyegddgtlctctvdftckqdgngcavkrcsqgkgtkvscscqkgykqgd 120

QY 984 YSCIEIDPCADSVNGSCHEHATCRMTGPCKHKCECKSHVGVGVDCEPRLPLDRCLQDN 1043  
DB 121 hscleldpaddlmggchelatckmtlpgkhkceckshvyvgdlncepeqjldlclqdn 180  
QY 1044 GOCHPDASCADLYPDDTMYGVFHLRSPGLGOYKLTFFDKAKCAKAPATATYNTQSTYAKK 1103  
DB 181 gqchadackcvdlhfgdtlvgrvfhltspjgqykltldkareacanaeatlnqjnsyagk 240

RESULT 9  
AAB42164  
ID AAB42164 standard. Protein: 330 AA.  
XX  
XX AAB42164;  
AC  
XX  
XX 08-FEB-2001 (first entry)  
DT  
XX  
XX Human ORFX ORF1928 polypeptide sequence SEQ ID NO:3856.  
DE  
XX  
XX Human; open reading frame; ORFX: detection; cytoskeletal; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;  
KW antineumatic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;  
KW thrombosis; contraceptive.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200058473-A2.  
PN  
XX  
XX 05-OCT-2000.  
PD  
XX  
XX 31-MAR-2000; 2000MO-US08621.  
PF  
XX  
XX 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX  
XX Shinkets RA, Leach M;  
PI  
XX  
XX WPI: 2000-602362/57.  
DR N-PSDB: AAC76373.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 11: Page 3007-3008; 5507pp: English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytoskeletal; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
CC antihypoid; and antineumatic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,



CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease: to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 330 AA;

	Query Match	14.1%	Score 1109.5:	DB 21:	Length 330:	
	Local Similarity	68.0%:	Pred No. 3.2e-67:			
	Matches 217:	Conservative 28:	Mismatches 63:	Indels 11:	Gaps 3	
OY	1089	AATATVYNQLSYAOKAKAYHLCASGMLIESGRVAVPTTYSAQKCAANYGVITDYGSRANKSE	1148			
Dd	2	aatmttttynqlsyayqakakyhlcsagwletgrvayptaftasqngcsgyvgivdvgyprpnkse	61			
OY	1149	MWDVCYEMKIDVNCCKKACGYVGDFGSCGNLLQVLMSFSLNFLTEVLAFSKSSARGQA	1208			
Dd	62	mwdvcyemkldvncckkacyvgdfgscgnllqvlimsfslnfltevlaysnsarsarga	121			
OY	1209	FLKLHTDISIRCTLEVPONSGLPNGKKSLSGRDIENHLTVNVNSFYNDLNGCFILRLMTLS	1268			
Dd	122	flehtldisirctltvlpqngsigmetltsrghlehlanvmnfnydlvngtkxparegk	181			
OY	1269	QLLTFSODQLHO--ETRVVDGRSIILOMDIITANGILHIISEPLRAPPPAAITAHSGLCT	1326			
Dd	182	qaahhcqpqrptxqplectrlvdgdralpqwlfsangilhispklkaparvllthglga	241			
OY	1337	GIFCAVVIVTGAIAIALAAYSFRLLKORTTGTFORFDOKRRTLMSSLVAS-----SSPRTSQ	1380			
Dd	242	gifcailllvgtavalaaysyfrlnrrrtlgfqbhfeseealnvaalqkqpenlsnplyest	301			
OY	1381	LCMRPQRHRPOSPPVTPSQ	1399			
Dd	302	tsapre---psydpftdse	317			
RESULT 10						
AAB83366	ID	AAB83366 standard; Protein: 334 AA.				
XX	AC	AAB83366;				
XX	DT	26-MAR-2002 (first entry)				
XX	DE	NOV10 protein sequence.				
XX	KW	NOV; Cytostatic; contractile; antiinflammatory; immunomodulatory;				
KW	KM	cardiovascular; casein kinase II phosphorylation site; contraction;				
KW	KM	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;				
KW	KM	epidermal growth factor; cell development; apoptosis; cell adhesion;				
KW	KM	growth migration; cell structure; motility; cancer; immune disorder;				
KW	KM	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;				
KW	KM	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;				
KW	KM	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;				
XX	XX	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.				
OS	Unidentified.					
XX	MO	WO200136638-A2.				
XX	PD	25-MAY-2001.				
XX	PF	17-NOV-2000; 2000WO-US31543.				
XX	PR	19-NOV-1999; 99US-166336P.				
XX	PR	29-NOV-1999; 99US-167785P.				
XX	PR	08-MAR-2000; 2000US-187844P.				
XX	PR	16-NOV-2000; 2000US-0715417.				

XX (CURA-) CURAGEN CORP.  
 PA  
 XX  
 XX Shimkets RA, Lichenstein H, Vernet C, Fernandes E;  
 PI  
 XX  
 DR WPI: 2001-648134/74.  
 DR N-PSDB; AAF87121.  
 XX  
 XX  
 PT Novel human polypeptides and the nucleic acids that encode them useful  
 PT for preventing, diagnosing and treating e.g. cancer, inflammation and  
 PT immune disorders -  
 XX  
 PS Claim 1; Page 35; 141pp: English.

CC This sequence is the NOV10 protein. The invention relates to  
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have  
CC Cytostatic; contractive; antiinflammatory; immunomodulatory; and  
CC cardiovascular activities. The sequences may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate NOX  
CC expression. They may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of protein by expressing inactive proteins or to  
CC supplement the patients own production of protein. They are used to  
CC produce NOX proteins, by inserting the nucleic acid into a cell and  
CC culturing it to express the protein. The DNA may be used as DNA probes in  
CC assays to detect and quantitate the presence of similar DNAs in samples,  
CC and which patients may need restorative therapy. The NOX protein may  
CC also be used as antigens in the production of antibodies (Abs) against  
CC NOX and in assays to identify modulators of NOX expression and  
CC activity. The anti-NOX Abs and antagonist are used to down regulate  
CC expression and activity. The anti-NOX Abs are used for detecting the  
CC presence of NOX in samples. Disorders that may be prevented, diagnosed  
CC and/or treated vary depending on the NOX protein. NOV1, NOV3, NOV5,  
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites  
CC characteristic of serine/threonine kinases, and are used to treat  
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular  
CC proliferation and contraction). NOV2-3, NOV6 and NOV8 are homologous to  
CC the epidermal growth factor (EGF)-like super family and are involved in,  
CC e.g., regulation of cell development, apoptosis, cell adhesion, growth  
CC migration, cell structure and motility and protein management, and are  
CC used to treat cancers, inflammatory disorders, immune disorders and  
CC cellular adhesion disorders. NOV9-10 are homologous to EGF-like  
CC fibrillin proteins and are used to treat cardiovascular disease e.g.  
CC hypertrophic cardiomyopathy, long-QT syndrome and Marfan syndrome.  
XX

[illegible]



Query Match 7.5%; Score 593.5; DB 11; Length 2189;  
 Best Local Similarity 24.1%; Pred. No. 4.5e-31;  
 Matches 293; Conservative 112; Mismatches 466; Indels 343; Gaps 71;

```

OY 132 IQKNBCNNNDITIVGEGCKCSQAP-----C-----PLETKPLRE--TRKCIYSIYPM 178
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 464 megnacsdid-----eceseasteipencvntegstfslækpyelvdgcv-kidf- 515
OY 179 GKRSVFICQPOCCVRTIITRACMLASLAHNAKPAPEYKMCALGTASWDVNGTGTCC 238
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 516 -----car-----gac-----nslahckenpeglaaic-----tcia 542
OY 239 GLGFMGTACETCTEGKYGHCDOACSV-----HGRCSGPGLGSGCDCCDVGWR-- 287
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 543 gysgdta-----gq-----hcdiddeciaendclpadggticentvgstckcaaygqd 593
OY 288 GVKCDMEITTDNC-NGT--CHTSANCLDPPDGKASCKACAGFRGNQVCTAINACETSN 344
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 594 gns-----fdidecangthnchasalc-Intgtsfecacnagfsgnvgendvdecstdad 649
OY 345 GCSTACADCKRTTPGNRVCCVCKAGY--TGDCIVCLEINPCLLENHGGCDRAECTOTGPNQ 403
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 650 dcgentlcnlv-gsflectcmagfeaadaclckdidecasgthctshactnta-gsft 707
OY 404 CNCLPKYTGDKVCSLINVCLTNNGCGSPFARCNTEEDORT-CTCKPDYTGDIIVCRGS 462
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 708 cecnpvgfdgdkcedvdcgqglhdcnvhaeeseadtlctctcgjaysqg----- 761
OY 463 IYGLPKNPSTSQYPFQLOEHAVRELACGPPTFAPLSSFNHEPRIKDWQOGLMSQV 522
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 762 -hge-----ngqgdidecaqdal-----cgentvctlnrpsfte-----cavc 797
OY 523 LRYHVVCCOQLLDNLKWTTSATTL-----QGEPVSISSODTYFINNEA----- 567
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 798 egfavvagk-----lkqatsltcididecndasknctacatsadgsccktagyscslp 851
OY 568 -----KVLSSDIISTNGVI--HVIDKLSPKNLLITPKDALGRVLOLTTVANHGVT 618
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 852 gfgqgdghctdidecatqgvgeha-----lceenagp--ynctlea--gyt 894
OY 619 KFSKLIDSGLSVITDSIHFPVTVPWPTDK-----ALEALPPEQODELFNODMND 669
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 895 -----gqdgavgcididecaastavlpnalcvntegsyftcecp-----gyrhcngct 944
OY 670 KIKSLTKFHVIRDSKALASDLPRASASWKTLOGSELSVRCGTGSD--IGELFLNEQMCRT 728
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 945 k-----idfctsekycnanasckenda-----glealctchsgyegnge--geegcknid 991
OY 729 RGLLFDCVAVGIDCLMNPFLTGRCDFPTTFDIPGEGSCIFTPKCPILKSKPKGVKKKC 788
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 992 -----ec-----svgepckdf-----gegvcdvdspsfiscscatgfk- 1025
OY 789 IYNPLPFRNVEGCONLCTVVIQTFRCHGVFMPDQACPGGPDTPCNNRGCRDLTYPM 848
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1026 -----frst--cgdi-----decldgk-mnlc--apvg-----gltctvtsf 1058
OY 849 GOCCLHTGTF--NGTACEILCMHGRFPCDQPRSCSEHGOCDEBIT-----GSEBCLCET 901
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1059 -tcscaagftgdlfctce-----didecataah-tcdpnatcvtvtsfegckegf 1107
OY 902 TAASCDTPFAVACTP---ACSVHATCTENN-----TCVCLNLYE--GCGITCTVWDPCK 952
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1108 sgdg--htcddidecadpnlkcdtlkylcngtsgtscgypyslaadgftcdhndeca 1166
OY 953 QNNGCAGVAVAKCSQKGTGVSCSKKGYKGDYSCEIIDPCADGVNGGCHAEHATCRMTG 1012
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1167 agtactgersfctvdqsgyckceckngyrgsgedcvddecceadvh-tsehatctnt-eg 1224
OY 1013 KHKCECKSHYVDGVDCPEBPLDLRCLQDNGCQHDPDASCADLYRQDPTTVGVFHLRSP 1072
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1225 shctcneqygqdgkckektvyp---c--dnspcgnamc-----eatadynctckag 1273

```

```

OY 1073 QXKLTFRKAEACAK-EAATIATYNOLSYAKAKYH-----LCSAGMLESGRAVPTTYA 1126
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1274 -----yemkdgaacvdidecsgtlnhcdphadcsntdgtctctcgsgyvgvlcedv---- 1325
OY 1127 SORCGANVVGIVDYSGSRANKSEMDVFCYRMKDVN-----CTCKAGYVGDGFSC 1175
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1326 -decaqnhag-----c-----dinavctnvpstfctcksgfegdghec 1363
OY 1176 SGNLLQVLMSPSLNFLTVEVLAFSKSSARG-----QAFKLHLNLDLSRGLFPQNSG 1229
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1364 tekvljlpqghdswlactcectaeatkqslrckvalplkxevhlcpldadisacqelgews 1423
OY 1230 LPG-NKSLSGRDIE 1242
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1424 cpyvdnlnshrrae 1437

```

RESULT 13  
 ABG06402  
 ID ABG06402 standard; Protein; 2912 AA.

XX ABG06402;  
 AC  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #6393.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS70589.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID NO 36761; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (II) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human





QY 180 KRSVFICOPQCVRTIIRACWLASIAHNAKPAPBEVKMCAIGTASV-----WD 228  
Db 327 -----atohdrvasifyc-----acpmgktgllchldaacvsnpchedaictdn 369  
QY 229 GYNGTGTCCGCGIGFNGTAC-----ETCTEKGKYGIRHCDQACSCYH--GRC--SQGPLDGDSD 281  
Db 370 pyngratctcpqgftgagcdqdvdecsly-----anpchehlyrcvntlygsf--lcq 418  
QY 282 CDVGWRGVKCDDEITTDNC--NGCTHTSANCILDPDGKASCKACAFRGNGTCTA--INAC 339  
Db 419 cgrlytgrctcdv--neclsgpctngatc--ldrtqfcticmagf--tglycevdidec 473  
QY 340 ETS---NGGCSSTRADCKRTTPGNRVCCVCKAGYTGDIYC--LEINPCLENHGGCDRNAECT 395  
Db 474 gspcvnggv-----ckdrvngfs--ctcpsgfts--tcqlvddecast--pcrnagkcv 523  
QY 396 QTEPNQAVCNCLPKRTGDKVCSLINVCLTNNGGCSFPAFCVTEODORI-----CTCKPD 451  
Db 524 dq-pdyecrcacgftg-----tlcdnvddcsdpd--chgyrcvtdgiasfscacp 573  
QY 452 YTGDDGIVCRGISYELPKNPSTSOYFFOLQEHAVRELAGPFTVAPL--SSSFNHEPRI 510  
Db 574 yrgtr--cesgv--decrspc-----rhgkcldlvdkylercpsgtgvcnevi 621  
QY 511 KDMDOO---GLMSQ--VLRYHVWGQQLL---LDNLKVTTSATTLQGEPPVSISVSDIYF 562  
Db 622 ddcasnpclfyrcrdginydcv--cqpgftgplcnveinecasspcgeg-----669  
QY 563 INNEAKVLSDDIISTNGVHVIDKLLSPKN---LLITPKDALGRVLQNLTVAAHNGYT 618  
Db 670 -----gscvdygengf-----rclcppgsjlpclclpshpcah-----epcshtg-- 707  
QY 619 KFSKLIDDS--GILSYITDSIHTPVTFWPTDKALEALPPEQODFLFNODNKKLKSYLK 676  
Db 708 ----lcydapggfrcvcepg-----wsgrpcsqsl-----733  
QY 677 FHVIRDSKALASDLPRSASWKTLQSELSVRCGTGSDIGELFLNEQMCRFIHRGLLFDVG 736  
Db 734 ----ard--acesqprcraqtcsdmgfhtcpgp-----763  
QY 737 VAVGIDCLLMNPPL-----GGRCDFFTFDIPGEGSCIFTPKPCPLKSPKGVK--KKCI 789  
Db 764 -vggrqcelispctnpcehgrces-----apqg-----lpvcscpqgwgprc-- 807  
QY 790 YNPLPRRNVGEC-----ONLCTVVIOTPRC--CH--GYFMPDCQACPGGPD--TPCNRK 838  
Db 808 -----qgdvdecaqaprcpgrphtcniagsfctchgytgspscddindcnpclng 861  
QY 839 GMCRD-----LYTPMG-----OCLCHTFNGTAC 862  
Db 862 gscqgdvgsfscscilpgfagrcardvdeclsnpcgpgtctdhvasfctcpcpgyggfnc 921  
QY 863 ELICWHRRFPDOPRSCSEHGOCDEGITSGBELCTGWTAAASC-----DTPAVFA 914  
Db 922 e-----qdlpdcspscfngtvcvgnv--nsfclcrpygvgancqtheadpcslrclng 975  
QY 915 VCTPA--CSYHATCTENNTVCNINLYEGDGLTC--TVVDFCKO---NNGGCAKVAKCSQKGT 969  
Db 976 vcsaahpgrfctclesft-----gpgcqlvdcwsrtpcpgng-----rcvqtga 1020  
QY 970 QVSCGCKKGTG-----DGYSCIEIDP-----CADGVNG 998  
Db 1021 y--clcpvgwsgfclclrsjprcreaaqjgvrleqlcga9gcvdedsshyvcvpegrtg 1078  
QY 999 G-----CHEHATCRMTGPGKHCKECKSHYVGDVGDCPEEQULPDRCLQDNQ 1045  
Db 1079 shbegevdpciaqpcqhggtcr--gymgymceclpgyngd--ncedd--vdecasq--p 1130  
QY 1046 CHPDASCADLYFQDTTGVGVHLSRPLGOYKLTFDKAKKAEAKKAAATATYNNQLSYAOKAK 1105  
Db 1131 cqhggsclclv-----arylscppgtlgyvlcelinedd--- 1163  
QY 1106 YHLCASAG--WLESG-----RVAVPTTASQKCGANVVGIVDYSRANKSE 1148

Db 1164 ---cgpplpdsqprclmngtclvvgfrcctcppyqglreadd-----ne 1208  
QY 1149 MWDVFCYR-----MKD-----VNCTCKAGYVG 1170  
Db 1209 crsgachaahtrdcldqdp9g9gfrclchagfsg 1240

Search completed: June 17, 2002, 12:30:21  
Job time: 252 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:30:24 ; Search time 16.67 Seconds  
(without alignments)  
3323.796 Million cell updates/sec

Title: US-09-842-930A-2  
Perfect score: 7861  
Sequence: 1 SLPSLLTRLEQMPDYSIIFRG.....WGHGCPDMRSQATVTVPR 1431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	510	6.5	2871	1 FBNI_MOUSE	O61554 mus musculu
2	508.5	6.5	2907	1 FBNI_MOUSE	O61555 mus musculu
3	497.5	6.3	2871	1 FBNI_HUMAN	P35555 homo sapien
4	496.5	6.3	2911	1 FBNI_HUMAN	P35556 homo sapien
5	486	6.2	2524	1 NOTC_XENLA	P21783 xenopus lae
6	484	6.2	2703	1 NOTC_DROME	P07207 drosophila
7	483.5	6.2	2871	1 FBNI_PIG	O9EV36 sus scrofa
8	482.5	6.1	2444	1 FBNI_HUMAN	P46531 homo sapien
9	481	6.1	2871	1 FBNI_BOVIN	P98133 bos taurus
10	473.5	6.0	2437	1 NOTC_BRARE	P46530 brachydantio
11	473	6.0	2318	1 NOTC_MOUSE	O61982 mus musculu
12	446	5.7	2531	1 NTCL_RAT	O07008 rattus norv
13	445	5.7	2531	1 NTCL_MOUSE	O01705 mus musculu
14	436.5	5.6	1964	1 NTCL_MOUSE	P31695 mus musculu
15	435	5.5	4289	1 TENX_HUMAN	P22105 homo sapien
16	416.5	5.3	1408	1 SERX_DROME	P81168 drosophila
17	413.5	5.3	1376	1 CRBH_HUMAN	P82279 homo sapien
18	403	5.1	1064	1 FBPI_STRPU	P10077 strongyloce
19	377.5	4.8	3110	1 LMA2_HUMAN	P24043 homo sapien
20	371	4.7	3051	1 YMX3_CAEEL	P34576 caenorhadi
21	357.5	4.5	1133	1 EGF_RAT	P07522 rattus norv
22	356	4.5	1799	1 LMB2_MOUSE	O61292 mus musculu
23	355.5	4.5	3106	1 LMB2_MOUSE	O60675 mus musculu
24	348.5	4.4	1786	1 LMB1_MOUSE	P02469 mus musculu
25	348.5	4.4	2139	1 CRB_DROME	P10040 drosophila
26	348	4.4	1801	1 LMB2_RAT	P15800 rattus norv
27	348	4.4	4544	1 LRPI_HUMAN	O07954 homo sapien
28	345	4.4	3695	1 LMA5_HUMAN	O15230 homo sapien
29	344	4.4	1786	1 LMB1_HUMAN	P07942 homo sapien
30	344	4.4	3084	1 LMA1_MOUSE	P19137 mus musculu
31	343	4.4	3075	1 LMA1_HUMAN	P25391 homo sapien
32	335	4.3	3672	1 LML2_CAEEL	O21313 caenorhadi
33	334	4.2	1790	1 LMB1_DROME	P11046 drosophila

34	333.5	4.2	1808	1 TENA_CHICK	P10039 gallus gall
35	332.5	4.2	1217	1 EGF_MOUSE	P01132 mus musculu
36	331	4.2	3712	1 LMA_DROME	O00174 drosophila
37	328	4.2	1207	1 EGF_HUMAN	P01133 homo sapien
38	328	4.2	1798	1 LMB2_HUMAN	P55268 homo sapien
39	326.5	4.2	1429	1 L112_CAEEL	P14585 caenorhadi
40	324.5	4.1	2201	1 TENA_HUMAN	P24821 homo sapien
41	324.5	4.1	4543	1 LRPI_CHICK	P98157 gallus gall
42	315.5	4.0	1247	1 NIDO_HUMAN	P14543 homo sapien
43	314.5	4.0	1746	1 NIDO_PIG	O29116 sus scrofa
44	313	4.0	1245	1 NIDO_MOUSE	P10493 mus musculu
45	301.5	3.8	3718	1 LMA5_MOUSE	O61001 mus musculu

## ALIGNMENTS

```

RESULT 1
FBNI_MOUSE STANDARD: PRT: 2871 AA.
AC O61554: 060826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Fibrillin 1 precursor.
OS FBNI OR FBNI-1.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; PubMed=7829516;
RA Yin W., Germiller J., Sanguineti C., Smiley E., Pangillan T.,
RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene."
RL J. Biol. Chem. 270:1798-1806(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC MICROFIBRILS (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L29454; AAA56840.1; -
DR EMBL: U22493; AAA64217.1; -
DR HSSP: P35555; IAPJ.
DR MGD: MGI:95469; Fbn1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR002212; TB.
DR Pfam: PF00683; TB; 9.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF_CA; 42.
DR SMART: SM00001; EGF_like; 4.

```

DR PROSITE; PS00010; ASX\_HYDROXYL; 43.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 38.  
DR PROSITE; PS01187; EGF\_CA; 45.  
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;  
Repeat; Signal; Multigene family.  
FT SIGNAL 1 27  
FT CHAIN 28 2871  
FT DOMAIN 81 112 EGF-LIKE 1, NON-CALCIUM BINDING.  
FT DOMAIN 115 146 EGF-LIKE 2, NON-CALCIUM BINDING.  
FT DOMAIN 147 178 EGF-LIKE 3, NON-CALCIUM BINDING.  
FT DOMAIN 246 287 EGF-LIKE 4, CALCIUM-BINDING.  
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM-BINDING.  
FT DOMAIN 330 401 TGFBR 1.  
FT DOMAIN 402 446 PRO-RICH.  
FT DOMAIN 449 489 EGF-LIKE 6, NON-CALCIUM BINDING.  
FT DOMAIN 490 529 EGF-LIKE 7, CALCIUM-BINDING.  
FT DOMAIN 530 571 EGF-LIKE 8, CALCIUM-BINDING.  
FT DOMAIN 572 612 EGF-LIKE 9, CALCIUM-BINDING.  
FT DOMAIN 613 653 EGF-LIKE 10, CALCIUM-BINDING.  
FT DOMAIN 656 721 TGFBR 2.  
FT DOMAIN 723 764 EGF-LIKE 11, CALCIUM-BINDING.  
FT DOMAIN 765 806 EGF-LIKE 12, CALCIUM-BINDING.  
FT DOMAIN 807 846 EGF-LIKE 13, CALCIUM-BINDING.  
FT DOMAIN 910 951 EGF-LIKE 14, CALCIUM-BINDING.  
FT DOMAIN 952 1018 TGFBR 3.  
FT DOMAIN 1028 1069 EGF-LIKE 15, CALCIUM-BINDING.  
FT DOMAIN 1070 1112 EGF-LIKE 16, CALCIUM-BINDING.  
FT DOMAIN 1113 1154 EGF-LIKE 17, CALCIUM-BINDING.  
FT DOMAIN 1155 1196 EGF-LIKE 18, CALCIUM-BINDING.  
FT DOMAIN 1197 1237 EGF-LIKE 19, CALCIUM-BINDING.  
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCIUM-BINDING.  
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCIUM-BINDING.  
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCIUM-BINDING.  
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCIUM-BINDING.  
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCIUM-BINDING.  
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCIUM-BINDING.  
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCIUM-BINDING.  
FT DOMAIN 1528 1559 TGFBR 4.  
FT DOMAIN 1560 1647 EGF-LIKE 27, CALCIUM-BINDING.  
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCIUM-BINDING.  
FT DOMAIN 1689 1758 TGFBR 5.  
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCIUM-BINDING.  
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCIUM-BINDING.  
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCIUM-BINDING.  
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCIUM-BINDING.  
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCIUM-BINDING.  
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCIUM-BINDING.  
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCIUM-BINDING.  
FT DOMAIN 2055 2121 TGFBR 6.  
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCIUM-BINDING.  
FT DOMAIN 2205 2246 EGF-LIKE 37, CALCIUM-BINDING.  
FT DOMAIN 2247 2290 EGF-LIKE 38, CALCIUM-BINDING.  
FT DOMAIN 2291 2332 EGF-LIKE 39, CALCIUM-BINDING.  
FT DOMAIN 2333 2400 EGF-LIKE 40, CALCIUM-BINDING.  
FT DOMAIN 2402 2443 TGFBR 7.  
FT DOMAIN 2444 2484 EGF-LIKE 41, CALCIUM-BINDING.  
FT DOMAIN 2485 2523 EGF-LIKE 42, CALCIUM-BINDING.  
FT DOMAIN 2524 2566 EGF-LIKE 43, CALCIUM-BINDING.  
FT DOMAIN 2567 2606 EGF-LIKE 44, CALCIUM-BINDING.  
FT DOMAIN 2607 2647 EGF-LIKE 45, CALCIUM-BINDING.  
FT DOMAIN 2648 2687 EGF-LIKE 46, CALCIUM-BINDING.  
FT DOMAIN 2688 2727 EGF-LIKE 47, CALCIUM-BINDING.  
FT DISULFID 85 94 BY SIMILARITY.  
FT DISULFID 89 100 BY SIMILARITY.  
FT DISULFID 102 111 BY SIMILARITY.  
FT DISULFID 119 129 BY SIMILARITY.  
FT DISULFID 123 134 BY SIMILARITY.  
FT DISULFID 136 145 BY SIMILARITY.  
FT DISULFID 150 160 BY SIMILARITY.  
FT DISULFID 154 166 BY SIMILARITY.  
FT DISULFID 168 177 BY SIMILARITY.  
FT DISULFID 250 262 BY SIMILARITY.

FT DISULFID 257 271 BY SIMILARITY.  
FT DISULFID 273 286 BY SIMILARITY.  
FT DISULFID 292 304 BY SIMILARITY.  
FT DISULFID 299 313 BY SIMILARITY.  
FT DISULFID 315 328 BY SIMILARITY.  
FT DISULFID 453 465 BY SIMILARITY.  
FT DISULFID 460 474 BY SIMILARITY.  
FT DISULFID 476 488 BY SIMILARITY.  
FT DISULFID 494 504 BY SIMILARITY.  
FT DISULFID 499 513 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 534 546 BY SIMILARITY.  
FT DISULFID 541 555 BY SIMILARITY.  
FT DISULFID 557 570 BY SIMILARITY.  
FT DISULFID 576 587 BY SIMILARITY.  
FT DISULFID 582 596 BY SIMILARITY.  
FT DISULFID 598 611 BY SIMILARITY.  
FT DISULFID 617 628 BY SIMILARITY.  
FT DISULFID 623 637 BY SIMILARITY.  
FT DISULFID 639 652 BY SIMILARITY.  
FT DISULFID 727 739 BY SIMILARITY.  
FT DISULFID 734 748 BY SIMILARITY.  
FT DISULFID 750 763 BY SIMILARITY.  
FT DISULFID 769 781 BY SIMILARITY.  
FT DISULFID 776 790 BY SIMILARITY.  
FT DISULFID 792 805 BY SIMILARITY.  
FT DISULFID 811 821 BY SIMILARITY.  
FT DISULFID 816 830 BY SIMILARITY.  
FT DISULFID 832 845 BY SIMILARITY.  
FT DISULFID 914 926 BY SIMILARITY.  
FT DISULFID 921 935 BY SIMILARITY.  
FT DISULFID 937 950 BY SIMILARITY.  
FT DISULFID 1032 1044 BY SIMILARITY.  
FT DISULFID 1039 1053 BY SIMILARITY.  
FT DISULFID 1055 1068 BY SIMILARITY.  
FT DISULFID 1074 1086 BY SIMILARITY.  
FT DISULFID 1081 1095 BY SIMILARITY.  
FT DISULFID 1097 1111 BY SIMILARITY.  
FT DISULFID 1117 1129 BY SIMILARITY.  
FT DISULFID 1124 1138 BY SIMILARITY.  
FT DISULFID 1140 1153 BY SIMILARITY.  
FT DISULFID 1159 1171 BY SIMILARITY.  
FT DISULFID 1166 1180 BY SIMILARITY.  
FT DISULFID 1182 1195 BY SIMILARITY.  
FT DISULFID 1201 1212 BY SIMILARITY.  
FT DISULFID 1208 1221 BY SIMILARITY.  
FT DISULFID 1223 1236 BY SIMILARITY.  
FT DISULFID 1242 1254 BY SIMILARITY.  
FT DISULFID 1249 1263 BY SIMILARITY.  
FT DISULFID 1265 1278 BY SIMILARITY.  
FT DISULFID 1284 1296 BY SIMILARITY.  
FT DISULFID 1291 1305 BY SIMILARITY.  
FT DISULFID 1307 1320 BY SIMILARITY.  
FT DISULFID 1326 1339 BY SIMILARITY.  
FT DISULFID 1333 1348 BY SIMILARITY.  
FT DISULFID 1350 1361 BY SIMILARITY.  
FT DISULFID 1367 1380 BY SIMILARITY.  
FT DISULFID 1374 1389 BY SIMILARITY.  
FT DISULFID 1391 1402 BY SIMILARITY.  
FT DISULFID 1408 1420 BY SIMILARITY.  
FT DISULFID 1415 1429 BY SIMILARITY.  
FT DISULFID 1431 1444 BY SIMILARITY.  
FT DISULFID 1450 1461 BY SIMILARITY.  
FT DISULFID 1456 1470 BY SIMILARITY.  
FT DISULFID 1472 1485 BY SIMILARITY.  
FT DISULFID 1491 1502 BY SIMILARITY.  
FT DISULFID 1497 1511 BY SIMILARITY.  
FT DISULFID 1513 1526 BY SIMILARITY.  
FT DISULFID 1610 1622 BY SIMILARITY.  
FT DISULFID 1617 1631 BY SIMILARITY.  
FT DISULFID 1633 1646 BY SIMILARITY.  
FT DISULFID 1652 1663 BY SIMILARITY.  
FT DISULFID 1658 1672 BY SIMILARITY.



CC	entries requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
DR	EMBL; L39790; AAA74908.1; -.		
DR	EMBL; S69359; AAC60685.1; -.		
DR	HSSP; P35555; IEMN.		
DR	MGI; MGI:95490; Fbn2.		
DR	InterPro; IPRO000152; Asx_hydroxyl.		
DR	InterPro; IPRO00561; EGF-like.		
DR	InterPro; IPRO01881; EGF_Ca.		
DR	InterPro; IPRO01438; EGF_II.		
DR	InterPro; IPRO02212; TB.		
DR	InterPro; IPRO08822; znf-C2H2.		
DR	Pfam; PF00008; EGF; 46.		
DR	Pfam; PF00683; TB; 9.		
DR	PRINTS; PR00010; EGFBLD.		
DR	SMART; SM00179; EGF_CA; 43.		
DR	SMART; SM00001; EGF_1like; 3.		
DR	PROSITE; PS00010; ASX_HYDROXYL; 43.		
DR	PROSITE; PS00022; EGF_1; 36.		
DR	PROSITE; PS01186; EGF_2; 36.		
DR	PROSITE; PS01187; EGF_CA; 43.		
KW	Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain; Repeat; Signal; Multigene family.		
KW	Repeat; Signal; Multigene family.		
FT	CHAIN	1	28
FT	DOMAIN	29	2907
FT	FT	111	142
FT	FT	145	176
FT	DOMAIN	176	208
FT	FT	276	317
FT	DOMAIN	318	359
FT	REPEAT	360	426
FT	DOMAIN	487	527
FT	DOMAIN	528	567
FT	DOMAIN	568	609
FT	DOMAIN	610	650
FT	DOMAIN	651	691
FT	REPEAT	692	760
FT	DOMAIN	761	802
FT	DOMAIN	803	844
FT	DOMAIN	845	883
FT	DOMAIN	948	989
FT	REPEAT	990	1065
FT	DOMAIN	1066	1107
FT	DOMAIN	1108	1150
FT	DOMAIN	1151	1192
FT	DOMAIN	1193	1234
FT	DOMAIN	1235	1275
FT	DOMAIN	1276	1317
FT	DOMAIN	1318	1359
FT	DOMAIN	1360	1400
FT	DOMAIN	1401	1441
FT	DOMAIN	1442	1483
FT	DOMAIN	1484	1524
FT	DOMAIN	1525	1565
FT	REPEAT	1566	1642
FT	DOMAIN	1643	1684
FT	DOMAIN	1685	1726
FT	REPEAT	1727	1800
FT	DOMAIN	1801	1842
FT	DOMAIN	1843	1884
FT	DOMAIN	1885	1926
FT	DOMAIN	1927	1965
FT	DOMAIN	1966	2008
FT	DOMAIN	2009	2048
FT	DOMAIN	2049	2090
FT	REPEAT	2091	2163
FT	DOMAIN	2164	2205
FT	DOMAIN	2206	2245
FT	DOMAIN	2246	2286
FT	DOMAIN	2287	2330
FT	DOMAIN	2331	2372
FT	REPEAT	2373	2441

FT	DOMAIN	2442	2483	EGF-LIKE 41	CALCIUM-BINDING.
FT	DOMAIN	2484	2524	EGF-LIKE 42	CALCIUM-BINDING.
FT	DOMAIN	2525	2563	EGF-LIKE 43	CALCIUM-BINDING.
FT	DOMAIN	2564	2606	EGF-LIKE 44	CALCIUM-BINDING.
FT	DOMAIN	2607	2646	EGF-LIKE 45	CALCIUM-BINDING.
FT	DOMAIN	2647	2687	EGF-LIKE 46	CALCIUM-BINDING.
FT	DOMAIN	2688	2727	EGF-LIKE 47	CALCIUM-BINDING.
FT	DISULFID	115	124	BY SIMILARITY.	
FT	DISULFID	119	130	BY SIMILARITY.	
FT	DISULFID	132	141	BY SIMILARITY.	
FT	DISULFID	149	159	BY SIMILARITY.	
FT	DISULFID	153	164	BY SIMILARITY.	
FT	DISULFID	166	175	BY SIMILARITY.	
FT	DISULFID	180	190	BY SIMILARITY.	
FT	DISULFID	184	196	BY SIMILARITY.	
FT	DISULFID	198	207	BY SIMILARITY.	
FT	DISULFID	280	292	BY SIMILARITY.	
FT	DISULFID	287	301	BY SIMILARITY.	
FT	DISULFID	303	316	BY SIMILARITY.	
FT	DISULFID	322	334	BY SIMILARITY.	
FT	DISULFID	329	343	BY SIMILARITY.	
FT	DISULFID	345	358	BY SIMILARITY.	
FT	DISULFID	491	503	BY SIMILARITY.	
FT	DISULFID	498	512	BY SIMILARITY.	
FT	DISULFID	514	526	BY SIMILARITY.	
FT	DISULFID	532	542	BY SIMILARITY.	
FT	DISULFID	537	551	BY SIMILARITY.	
FT	DISULFID	553	566	BY SIMILARITY.	
FT	DISULFID	572	584	BY SIMILARITY.	
FT	DISULFID	579	593	BY SIMILARITY.	
FT	DISULFID	595	608	BY SIMILARITY.	
FT	DISULFID	614	625	BY SIMILARITY.	
FT	DISULFID	620	634	BY SIMILARITY.	
FT	DISULFID	636	649	BY SIMILARITY.	
FT	DISULFID	655	666	BY SIMILARITY.	
FT	DISULFID	661	675	BY SIMILARITY.	
FT	DISULFID	677	690	BY SIMILARITY.	
FT	DISULFID	765	777	BY SIMILARITY.	
FT	DISULFID	772	786	BY SIMILARITY.	
FT	DISULFID	788	801	BY SIMILARITY.	
FT	DISULFID	807	819	BY SIMILARITY.	
FT	DISULFID	814	828	BY SIMILARITY.	
FT	DISULFID	830	843	BY SIMILARITY.	
FT	DISULFID	849	859	BY SIMILARITY.	
FT	DISOLFTD	854	868	BY SIMILARITY.	
FT	DISOLFTD	870	883	BY SIMILARITY.	
FT	DISOLFTD	952	964	BY SIMILARITY.	
FT	DISOLFTD	959	973	BY SIMILARITY.	
FT	DISOLFTD	975	988	BY SIMILARITY.	
FT	DISOLFTD	1070	1082	BY SIMILARITY.	
FT	DISOLFTD	1077	1091	BY SIMILARITY.	
FT	DISOLFTD	1093	1106	BY SIMILARITY.	
FT	DISOLFTD	1112	1124	BY SIMILARITY.	
FT	DISOLFTD	1119	1133	BY SIMILARITY.	
FT	DISOLFTD	1135	1149	BY SIMILARITY.	
FT	DISOLFTD	1155	1167	BY SIMILARITY.	
FT	DISOLFTD	1162	1176	BY SIMILARITY.	
FT	DISOLFTD	1178	1191	BY SIMILARITY.	
FT	DISOLFTD	1197	1209	BY SIMILARITY.	
FT	DISOLFTD	1204	1218	BY SIMILARITY.	
FT	DISOLFTD	1220	1233	BY SIMILARITY.	
FT	DISOLFTD	1239	1250	BY SIMILARITY.	
FT	DISOLFTD	1246	1259	BY SIMILARITY.	
FT	DISOLFTD	1261	1274	BY SIMILARITY.	
FT	DISOLFTD	1280	1292	BY SIMILARITY.	
FT	DISOLFTD	1287	1301	BY SIMILARITY.	
FT	DISOLFTD	1303	1316	BY SIMILARITY.	
FT	DISOLFTD	1322	1334	BY SIMILARITY.	
FT	DISOLFTD	1329	1343	BY SIMILARITY.	
FT	DISOLFTD	1345	1358	BY SIMILARITY.	
FT	DISOLFTD	1364	1377	BY SIMILARITY.	
FT	DISOLFTD	1371	1386	BY SIMILARITY.	
FT	DISOLFTD	1388	1399	BY SIMILARITY.	

```

FT DISULFID 1405 1418 BY SIMILARITY.
FT DISULFID 1412 1427 BY SIMILARITY.
FT DISULFID 1429 1440 BY SIMILARITY.
FT DISULFID 1446 1458 BY SIMILARITY.
FT DISULFID 1453 1467 BY SIMILARITY.
FT DISULFID 1469 1482 BY SIMILARITY.
FT DISULFID 1488 1499 BY SIMILARITY.
FT DISULFID 1494 1508 BY SIMILARITY.
FT DISULFID 1510 1523 BY SIMILARITY.
FT DISULFID 1529 1540 BY SIMILARITY.
FT DISULFID 1535 1549 BY SIMILARITY.
FT DISULFID 1551 1564 BY SIMILARITY.
FT DISULFID 1647 1659 BY SIMILARITY.
FT DISULFID 1654 1668 BY SIMILARITY.
FT DISULFID 1670 1683 BY SIMILARITY.
FT DISULFID 1689 1701 BY SIMILARITY.

Query Match 6.5%, Score 508.5; DB 1; Length 2907;
Best Local Similarity 23.5%; Pred. No. 1,3e-27;
Matches 296; Conservative 123; Mismatches 395; Indels 445; Gaps 87;

QY 190 QCVRIITRAQMLASLANAKPAPG---EVKMCALGTASVMDGVNGTGT--CQCGLGFGNG 244
DB 502 KCIPTVSSYRC-RCNMGYK-QDANGDCIDVDECTSNPCSNQCVNTPSSYYCKCHAGFOR 559
QY 245 T---ACEETGEGKXGHCDO-ACSCVHGRCSOGPFGDGS--CDGDVGWGVKCDMEITTT 297
DB 560 TPTKQACIDIDE-----CIQNGVYCKNGRCVN---SDGSPQICINAGF-----ELFTT 603
QY 298 D--NC--NGTHTSANCIL---LDPDGKASCKCAAGF--RGNGTVCIAINACT-----SN 343
DB 604 DGNKNCVDHDECTTTNMCLNGMCINEDGSFKCYCKPGFLLAPNGRYCTDVEDCQTPGICMN 663
QY 344 GGC-----STRADCKRTTPG-----NRVCV----- 363
DB 664 GHCIINNEGSPFDCD---PRLGAVGVDRVCYDTNMRSTCYGEIKKGVCPVREPAGAVTKSE 720
QY 364 ----- 363
DB 721 CCCANPDYGFGEPCOPCPAKNSAEFHGLCSSGITTVGGRDINECALDPDICANGICENTL 780
QY 364 -----CKAGYTG D--GIYCLEINPCLENNHGGCDRNAECTOTGPNQAVCNCLPKRYT--G 412
DB 781 RGYTRCNCNSGEYBPASGRNCIDIDECLVNRLLCD--NGLCRNT--PGSYSCCTCPGCVLPT 838
QY 413 DGKVCSLINVLCTN---NGGSGPFAFCVNTBDDQRICTCKP--DYTGGIYVCRGSITYG-- 465
DB 839 ETECEDVNECESNFCVNGACR-----NNLGSFHCECSPSKSLSTGLICIDISLAKTIC 891
QY 466 -----ELPKNPSTQYFFOLOEHAHVRELACGPFTVFAPLSSSNHPRIDMDQOQ 516
DB 892 WLNIODNCEVNINGAT-----LKSECCATLIGA-----WQSP 924
QY 517 GIMSOVLRYHVVGCQQLLDNLKVTTSATTLQGEPVSVISODIVEINNEAKVL-----S 571
DB 925 -----CERCELD-----AACPRGFARIKGVTCEDV--NECEVFPFGVCPN 961
QY 572 SDIISTNGVIVHIDKLSPKULLITPKALGRVLONLTTVAANHGYTFPSKLIODSGLIS 631
DB 962 GRCVNSKSKSFH---CECPBELTLL--DGTGRVCLD---IRMEHCFLEKWD--EDECIHP 1008
QY 632 V-----ITDSIHTPYTVFMPD-----KALEALPREQODFLFNQONKTKLYLKF- 677
DB 1009 VPGAFRMDACCAGVAMAGTCEBCKPKGTKEVEFTLCPRGFAFA---NRGDILTRGPRFC 1064
QY 678 HVIRDSKALASDPRASWKTLOGSELSVRCGTG-----SDIGELFLNEQMCRFI 727
DB 1065 KDIHECKAFPCMCYTGKCRNTI--GSFKRCNNNGFALDMEERNCTDIDECRISPLDC--- 1119
QY 728 HRGLLPDVGAVYGDIC-----LLM-----NPTL--GGRC-DPFTTFDIDIGE 765
DB 1120 GSGICVNTPGSFCECEGEGYESGFMNMKNCDIDECERNPLLCRGTCVNTGSGF-----Q 1175

```

```

QY 766 GCSGIEFTKCPKPSKPKGVKKKCI--YNPLPRRRNVEGCON-LCTVVIQTTPRC-CH-GY-F 820
DB 1176 C-----DCPLIGHELSPSREDCVDIDNCSLSDNL--CENKGCVMYMIIGTYQSCNPGYCA 1226
QY 821 MPDQACP-----GGPDTPCNNRGMCRDLVTPRMGQCLCHTGP---NGTAC---EL 864
DB 1227 TPDROGCTDIDECMIIMGCDTQCTN---SEGSY-----ECSCSEGYALMPDGRSCADIDE 1279
QY 865 CHHGFGRPDQORRSSEHQCDGEGITGSGECLCEGTWTAASCDTTPAV-FAVG--TPRACS 921
DB 1280 C-----ENNDPIC-DGGQC--TNIPGEYRCLCYDGF-MASMDMTCTCIDVNECDLNPNIC 1329
QY 922 VVATCTENN-----TCVCNLNVE--GDGIT-CTVVDPCQKONGGCACAVAKSKQGTQVSCG 975
DB 1330 MEGEC-ENITKSGFTHCQGLGYSVKKGTCTTCTVDECEICAHNCMDHNASLNPVGSFKSCG 1388
QY 976 KKGYKQDGYSCIEIDPCADGVNGGCHENATCRMTGPKKHCKECSKSHYVGDVDECEPDQL 1035
DB 1389 REGWNGNGIKCIDLDEBCANGTH-QCSINAQCVNT--PGSYRCACSEGFQDGFQTS----D 1442
QY 1036 LDRC-----LDNNGOC-----HPDASCADL---YQDITV----- 1062
DB 1443 VDECAENTNLCENGQCLNVPNAVRCCEEMGFPPASDSRSCQDIDECSPQNTICVFQTCNNL 1502
QY 1063 -GVFHLRSPLOGYKLTFPKAKEACA---KEATTIATYNOLSYAOKAKYHLCASAGULESGR 1118
DB 1503 PGMFHCICDDG-YEL--DRTGNCNTDIDECADPINCYNGLCVNTGRIE-CMC----- 1551
QY 1119 VAVPTTYASQKCGANVGVIVDYGSRANKSEMDVFCYRMKMDVNTCKAKAGVGDG-FSCS 1176
DB 1552 ---PPDF---QLNATGVGVD--NRVG-----NCYLFGPRDGSLSGN 1587

RESULT 3
FBN1_HUMAN
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93372860; PubMed=8364578;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Panglilian T., Bonadio J.;
RT "Genomic organization of the sequence coding for fibrillin, the
RT defective gene product in Marfan syndrome.";
RL Hum. Mol. Genet. 2:961-968(1993).
[2]
RP SEQUENCE OF 1-932 FROM N.A.
RC TISSUE=Placenta; and Fibroblast;
RX MEDLINE=94010947; PubMed=7691719;
RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT multidomain structure and alternatively spliced exons at the 5'
RT end.";
RL Genomics 17:476-484(1993).
[3]
RP SEQUENCE OF 899-2871 FROM N.A.
RX MEDLINE=91304568; PubMed=1852207;
RA Madlen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RT "Partial sequence of a candidate gene for the Marfan syndrome.";
RL Nature 352:334-337(1991).
[4]
RP SEQUENCE OF 813-1313 FROM N.A.
RX MEDLINE=91304567; PubMed=1852206;

```

RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,  
 RA Tsipouras P., Ramirez F., Hollister D.W.;  
 RT "Linkage of Marfan syndrome and a phenotypically related disorder to  
 RL two different fibrillin genes."; *Nature* 352:330-334(1991).  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=91317849; PubMed=1860873;  
 RA Sakai L.Y., Keene D.R., Ghanville R.W., Bachinger H.P.;  
 RT "Purification and partial characterization of fibrillin, a cysteine-  
 RT rich structural component of connective tissue microfibrils."; *J. Biol. Chem.* 266:14763-14770(1991).  
 RL [6]  
 RN STRUCTURE BY NMR OF 2054-2125.  
 RX MEDLINE=98031893; PubMed=9362480;  
 RA Yuan X., Downing A.K., Knott V., Handford P.A.;  
 RT "Solution structure of the transforming growth factor beta-binding  
 RT protein-like module, a domain associated with matrix fibrils."; *EMBO J.* 16:6659-6666(1997).  
 RL [7]  
 RN STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=96144829; PubMed=8568869;  
 RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;  
 RT "Calcium binding properties of an epidermal growth factor-like domain  
 RT pair from human fibrillin-1."; *J. Mol. Biol.* 255:22-27(1996).  
 RL [8]  
 RN STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=96222301; PubMed=8653794;  
 RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,  
 RA Handford P.A.;  
 RT "Solution structure of a pair of calcium-binding epidermal growth  
 RT factor-like domains: implications for the Marfan syndrome and other  
 RT genetic disorders."; *Cell* 85:597-605(1996).  
 RL [9]  
 RN REVIEW ON MFS VARIANTS.  
 RX MEDLINE=96174615; PubMed=8594563;  
 RA Colod-Beroud G., Beroud C., Soussi T., Junten C., Boileau C.;  
 RT "Software and database for the analysis of mutations in the human  
 RT FBN1 gene."; *Nucleic Acids Res.* 24:137-141(1996).  
 RL [10]  
 RN REVIEW ON MFS VARIANTS.  
 RX MEDLINE=97169383; PubMed=9016526;  
 RA Colod-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,  
 RA Godfrey M., Hayward C., Karttunen L., Mllewicz D., Peltonen L.,  
 RA Richards R.J., Wang W., Junten C., Boileau C.;  
 RT "Marfan Database (second edition): software and database for the  
 RT analysis of mutations in the human FBN1 gene."; *Nucleic Acids Res.* 25:147-150(1997).  
 RL [11]  
 RN REVIEW ON VARIANTS.  
 RP MEDLINE=98062175; PubMed=9401003;  
 RA Hayward C., Brock D.J.H.;  
 RT "Fibrillin-1 mutations in Marfan syndrome and other type-1  
 RT fibrillinopathies."; *Hum. Mutat.* 10:415-423(1997).  
 RL [12]  
 RN VARIANT MFS PRO-1137.  
 RP MEDLINE=91304569; PubMed=1852208;  
 RX Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,  
 RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,  
 RA Curstlin S.M., Stetten G., Meyers D.A., Francomano C.A.;  
 RT "Marfan syndrome caused by a recurrent de novo missense mutation in  
 RT the fibrillin gene."; *Nature* 352:337-339(1991).  
 RL [13]  
 RN VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.  
 RP MEDLINE=93250834; PubMed=1301946;  
 RX Dietz H.C., Saraiya J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;  
 RT "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome  
 RT patients at cysteine residues in EGF-like domains.";

RL Hum. Mutat. 1:366-374(1992).  
 RN [14]  
 RP VARIANT MFS SER-2307.  
 RX MEDLINE=92235290; PubMed=1569206;  
 RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendior R.J. Jr.,  
 RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;  
 RT "Marfan phenotype variability in a family segregating a missense  
 RT mutation in the epidermal growth factor-like motif of the fibrillin  
 RT gene."; *J. Clin. Invest.* 89:1674-1680(1992).  
 RL [15]  
 RN VARIANTS MFS ILE-548 AND ALA-723.  
 RX MEDLINE=94010946; PubMed=8406497;  
 RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,  
 RA Pyeritz R.E., Francomano C.A.;  
 RT "Four novel FBN1 mutations: significance for mutant transcript level  
 RT and EGF-like domain calcium binding in the pathogenesis of Marfan  
 RT syndrome."; *Hum. Mol. Genet.* 17:468-475(1993).  
 RL [16]  
 RN VARIANTS MFS SER-2144.  
 RX MEDLINE=93278402; PubMed=8504310;  
 RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;  
 RT "A novel fibrillin mutation in the Marfan syndrome which could  
 RT disrupt calcium binding of the epidermal growth factor-like module."; *Hum. Mol. Genet.* 2:475-477(1993).  
 RL [17]  
 RN VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.  
 RX MEDLINE=94108431; PubMed=8281141;  
 RA Tyan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,  
 RA Berg M.A., Miller D.C., Francke U.;  
 RT "Mutation screening of complete fibrillin-1 coding sequence: report  
 RT of five new mutations, including two in 8-cysteine domains."; *Hum. Mol. Genet.* 2:1813-1821(1993).  
 RL [18]  
 RN VARIANTS MFS GLY-217 AND ARG-2627.  
 RX MEDLINE=95067970; PubMed=797366;  
 RA Karttunen L., Raghunath M., Loenqvist L., Peltonen L.;  
 RT "A compound-heterozygous Marfan patient: two defective fibrillin  
 RT alleles result in a lethal phenotype."; *Am. J. Hum. Genet.* 55:1083-1091(1994).  
 RL [19]  
 RN VARIANT EL LYS-2447.  
 RX MEDLINE=94245249; PubMed=8186302;  
 RA Longqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,  
 RA Peltonen L.;  
 RT "A novel mutation of the fibrillin gene causing ectopia lentis."; *Genomics* 19:573-576(1994).  
 RL [20]  
 RN VARIANT MFS CYS-627.  
 RX MEDLINE=94272487; PubMed=8004112;  
 RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;  
 RT "Two novel mutations and a neutral polymorphism in EGF-like domains  
 RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan  
 RT syndrome patients."; *Hum. Mol. Genet.* 3:373-375(1994).  
 RL [21]  
 RN VARIANT MFS CYS-122.  
 RX MEDLINE=94314977; PubMed=8040326;  
 RA Stahl-Hallengren C., Urkonen T., Kainulainen K., Kristofersson U.,  
 RA Saxne T., Tornqvist K., Peltonen L.;  
 RT "An extra cysteine in one of the non-calcium-binding epidermal growth  
 RT factor-like motifs of the FBN1 polypeptide is connected to a novel  
 RT variant of Marfan syndrome."; *J. Clin. Invest.* 94:709-713(1994).  
 RL [22]  
 RN VARIANT MFS TYR-1223.  
 RX MEDLINE=94351682; PubMed=8071963;  
 RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;  
 RT "A new missense mutation of fibrillin in a patient with Marfan  
 RT syndrome."; *J. Med. Genet.* 31:338-339(1994).  
 RL [23]

RP VARIANT MFS HIS-1170.  
 RX MEDLINE=95174777; PubMed=7870075;  
 RA Hayward C., Porteous M.E.M., Brock D.J.H.;  
 RT "A novel mutation in the fibrillin gene (FBN1) in familial  
 RT arachnodactyly.";  
 RL Mol. Cell. Probes 8:325-327(1994).  
 RN [24]  
 RP VARIANTS MFS G-217; N-1023; R-1074; Y-1242; R-1513; E-2127; W-2151;  
 RP K-2447 AND R-2511.  
 RX MEDLINE=94184368; PubMed=8136837;  
 RA Kainulainen K., Karttunen L., Punakka L., Sakai L., Peltonen L.;  
 RT "Mutations in the fibrillin gene responsible for dominant ectopia  
 RT lentis and neonatal Marfan syndrome.";  
 RL Nat. Genet. 6:64-69(1994).  
 RN [25]  
 RP VARIANT SER-1127.

Query Match 6.3%; Score 497.5; DB 1; Length 2871;  
 Best Local Similarity 21.7%; Pred. No. 8e-27;  
 Matches 261; Conservative 121; Mismatches 356; Indels 465; Gaps 70;

QY 189 POCRTITTRACWLAS-LAHNAK--PARG-----EVMKCALGTASV 226  
 DB 442 PRVLPLVNTDVCQLVRYLCQNGRCIPTPGYSRCECNKGFOLDRECEIDVECEKKNPCAG 501  
 QY 227 WDGVNGTG--TCQCGLGNFACETCTEGKYGHCDO-ACSCVHRCGSGPLDGS--CD 281  
 DB 502 GECINNGSYTCQCRAGYOSTL--TRTECRDIDECLONGRICNNRCIN--TDGSPHCY 556  
 QY 282 CDVGMR---GVKCDMEITTDNC-----NGCHTSANCLDPDGKASCKACAGFR--G 328  
 DB 557 CNAGFHVTRDGNCE--DMDECSLRNMLNGMC-----INDEGSKICCKKPEFOLAS 606  
 QY 329 NGYCTAINACET--SNGGC-----KRTTPGNRY-- 361  
 DB 607 DGRYCKDINECEKRCVNTGSGYRCEPGLAVGLDGRVCVDTNMRSTCYGK 666  
 QY 347 -----STRADC-----KRTTPGNRY-- 361  
 DB 667 RGCCIKPLFGAVTSECCCASTEYAFGEPCOPCPAONSAEYQALSSGPGMTSAGSDINE 726  
 QY 362 -----CVCKAGYTG--GIVCLEINPCLENHGGCDRNAECTQTG 398  
 DB 727 CALDPDLCPNGICENTLRGTAKCICNSGYEVDSTGKNCVDINCEVNLNLCD-NGCCRNLT- 784  
 QY 339 PNOAVCNKLPK---YTGAGKVCSLINVCLTN--NGGC--SPFAFCNTTEDDQRTCTKP 450  
 DB 785 PGSEFCTC-PRGFIYKPLKTCEDIDECSSPCINGVCNKSPPSP-----ICECSS 834  
 QY 451 DYTGD--GIVCRGSYVGLPKNPSTSYQFQLOEHAVALAGPGEFTVAFPLSSSFNHP 508  
 DB 835 ESTIDPTTICETIKGT----- 852  
 QY 509 RIKMDQGLMSQYLRYHVAGCQQLLD-NLKVTTSATTLQ-----GEFVSISV 556  
 DB 853 -----CMQTVIDGRCEININGATLKSCCCSSGAAANGSPCTL-C 890  
 QY 557 SODTVFINNEAKVLS--DI-----ISTNGVIHVIDLSPKNLLITPK 598  
 DB 891 QVDPICGKYSRIKGTQCEDIDEEVEFPGVCKNGLCVNTRSF--KCOCPSGMTL--- 943  
 QY 599 DALGRVIONLTVAANHGVTKFSKLIDSGLLSVI--TDSIHTPVTFVMTDKA---- 650  
 DB 944 DATGRICIDIL-----ETCFRLREDECHLPIAGRRHMDACCSVCAANGTECECECP 997  
 QY 651 -----LEALPEQODFLFNODNKKDKLSYLKFHVIRDSKALASDLPRASAKWTLQSGEL 704  
 DB 998 MRNTPYEELCPRGPGA---TKEITNGKPEFFKIDNECKMIPSLCTHGKCKNTI--GSF 1051  
 QY 705 SVRCGTG-----SDIGELFLNEQMKRFIHHGLF-----DGVAVG---- 740  
 DB 1052 KCRDSDGALDSEBRNCTDIDECRISPDLC---GRGQCVNTPGDECKCDEGYESGFMM 1108

QY 741 -----IDCLLNPTL--GG-----RCDFTTFDIPGEGSCIFTPKPLSK--PK 782  
 DB 1109 KNMCIDECQDRPLLCRGVCHNTEGSRCEBPRPHOLSPNISACIDINEELSHLCRN 1168  
 QY 783 G-----VKKKCIYNP-----LP--FRNVE-----GCONLCTVVIQTTPRC-CHGYF- 820  
 DB 1169 GRCVNLIGKYOCACNPGYHSTPDRFLFCVIDECSIMNGCEFTNSEGSEYSCOPFA 1228  
 QY 821 -MPDQACRG--GPDTP--CNRMCMCDL---YTPMOGCIHCHGFNCTACSLCHNHRFG 871  
 DB 1229 LMPDQRSTDIDECEDNPNDICG-GQCTNIPGEY---RCLCYDGFMAE----- 1273  
 QY 872 PDQPRSCSEHGQCD-----EGITSGECLCEGTWTAASCDPTPAFAVC--T 917  
 DB 1274 ---DMKTCVDVNECDNLNICSSTGCEMTKSFICHCMDGYSKKKRGCGDINCELGA 1330  
 QY 918 PACSVHANCETENN--CVCNLNVEBGDITTCYVDFCKONNGSCARVAKSOKGTQVS 972  
 DB 1331 HNCGRHAVCT--NTAGSEKSCSPGIMGDIKCTDIDECSSMCTHMSOHADCKNTMGSYR 1388  
 QY 973 CSCKKGYKGDGSCLEIDPCADGV-----NGCCHENATGRMTPGRKHKCEKSHYV--GDG 1026  
 DB 1389 CLCKRGYTGDDFTCTDIDECSENLMLCGNGQC-----LNAFGYRCECDMGFVPSADG 1441  
 QY 1027 VDCEPEQLPLDRCLQDN---GQCHPDASCADLYPDFTVGVFHLRSPLGQYKTLFDRK 1082  
 DB 1442 KACE-----DIDECSLPNICVGTCH-----NLPLGLRCECEIG-YEL--DRSG 1482  
 QY 1083 EAC 1085  
 DB 1483 GNC 1485

RESULT 4  
 FBN2\_HUMAN  
 ID FBN2\_HUMAN STANDARD; PRT; 2911 AA.  
 AC P35556;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Fibrillin 2 precursor.  
 GN FBN2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RX MEDLINE=94165105; PubMed=8120105;  
 RA Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguinetti C.,  
 RA Bonadio J., Mecham R.P., Ramirez F.;  
 RT "Structure and expression of fibrillin-2, a novel microfibrillar  
 RT component preferentially located in elastic matrices.";  
 RL J. Cell Biol. 124:855-863(1994).  
 RN [12]  
 RP SEQUENCE OF 752-1505 FROM N.A.  
 RX MEDLINE=91304567; PubMed=1852206;  
 RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,  
 RA Tsipouras P., Ramirez F., Hollister D.;  
 RT "Linkage of Marfan syndrome and a phenotypically related disorder to  
 RT two different fibrillin genes.";  
 RL Nature 352:330-334(1991).  
 RN [13]  
 RP VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.  
 RX MEDLINE=96083599; PubMed=7493032;  
 RA Punam E.A., Zhang H., Ramirez F., Milewicz D.M.;  
 RT "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,  
 RT congenital contractual arachnodactyly.";  
 RL Nat. Genet. 11:456-458(1995).  
 RN [14]  
 RP VARIANTS CCA HIS-1114.  
 RX MEDLINE=98407789; PubMed=9737771;  
 RA Babcock D., Gasner C., Francke U., Maslen C.;

RT "A single mutation that results in an asp-to-his substitution and  
RT partial exon skipping in a family with congenital contractual  
RT arachnoidectyly.";  
RL Hum. Genet. 103:22-28(1998).  
RN [5]  
RP VARIANTS CCA PHE-1141 AND TRP-1252.  
RX MEDLINE=20259236; Pubmed=10797416;  
RA Belien S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,  
RA Godfrey M.;  
RT "Two novel fibrillin-2 mutations in congenital contractual  
RT arachnoidectyly.";  
RL Am. J. Med. Genet. 92:7-12(2000).  
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
CC THAT BINDS CALCIUM. FIBRILLIN-2-CONTAINING MICROFIBRILS REGULATE  
CC THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.  
CC -1- DISEASE: DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTURAL  
CC ARACHNOECTYLY (CCA) (ALSO KNOWN AS BEALS SYNDROME). CCA IS  
CC PHENOTYPICALLY SIMILAR TO MARFAN SYNDROME, BUT DOES NOT EFFECT THE  
CC AORTA AND THE EYES.  
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING  
CC EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U03272; AAA18950.1; -;  
DR EMBL; X62009; -; NOT\_ANNOTATED\_CDS.  
DR PIR; S17063; S17063.  
DR PIR; S31101; S31101.  
DR HSSP; P35555; IEMN.  
DR MIM; 121050; -;  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF\_II.  
DR InterPro: IPR002212; TB.  
DR Pfam; PF00008; EGF; 45.  
DR Pfam; PF00683; TB; 9.  
DR PRINTS; PR00010; EGFBL00D.  
DR SMART; SM00179; EGF\_CA; 43.  
DR SMART; SM00001; EGF\_like; 3.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 43.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 37.  
DR PROSITE; PS01187; EGF\_CA; 43.  
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;  
KW Repeat; Signal; Multigene family; Disease mutation; Polymorphism.  
FT SIGNAL 1 28  
FT CHAIN 29 2911  
FT DOMAIN 111 142 EGF-LIKE 1, NON-CALCIUM BINDING.  
FT DOMAIN 145 176 EGF-LIKE 2, NON-CALCIUM BINDING.  
FT DOMAIN 176 207 EGF-LIKE 3, NON-CALCIUM BINDING.  
FT DOMAIN 275 316 EGF-LIKE 4, CALCIUM-BINDING.  
FT DOMAIN 317 358 EGF-LIKE 5, CALCIUM-BINDING.  
FT REPEAT 359 425 TGFBR 1.  
FT DOMAIN 493 533 EGF-LIKE 6, NON-CALCIUM BINDING.  
FT DOMAIN 534 573 EGF-LIKE 7, CALCIUM-BINDING.  
FT DOMAIN 574 615 EGF-LIKE 8, CALCIUM-BINDING.  
FT DOMAIN 616 656 EGF-LIKE 9, CALCIUM-BINDING.  
FT DOMAIN 657 697 EGF-LIKE 10, CALCIUM-BINDING.  
FT REPEAT 698 766 TGFBR 2.  
FT REPEAT 767 808 EGF-LIKE 11, CALCIUM-BINDING.  
FT DOMAIN 809 850 EGF-LIKE 12, CALCIUM-BINDING.  
FT DOMAIN 851 890 EGF-LIKE 13, CALCIUM-BINDING.  
FT DOMAIN 954 995 EGF-LIKE 14, CALCIUM-BINDING.  
FT REPEAT 996 1071 TGFBR 3.  
FT DOMAIN 1072 1113 EGF-LIKE 15, CALCIUM-BINDING.

FT DOMAIN 1114 1156 EGF-LIKE 16, CALCIUM-BINDING.  
FT DOMAIN 1157 1198 EGF-LIKE 17, CALCIUM-BINDING.  
FT DOMAIN 1199 1240 EGF-LIKE 18, CALCIUM-BINDING.  
FT DOMAIN 1241 1281 EGF-LIKE 19, CALCIUM-BINDING.  
FT DOMAIN 1282 1323 EGF-LIKE 20, CALCIUM-BINDING.  
FT DOMAIN 1324 1365 EGF-LIKE 21, CALCIUM-BINDING.  
FT DOMAIN 1366 1406 EGF-LIKE 22, CALCIUM-BINDING.  
FT DOMAIN 1407 1447 EGF-LIKE 23, CALCIUM-BINDING.  
FT DOMAIN 1448 1489 EGF-LIKE 24, CALCIUM-BINDING.  
FT DOMAIN 1490 1530 EGF-LIKE 25, CALCIUM-BINDING.  
FT DOMAIN 1531 1571 EGF-LIKE 26, CALCIUM-BINDING.  
FT REPEAT 1572 1648 TGFBR 4.  
FT DOMAIN 1649 1690 EGF-LIKE 27, CALCIUM-BINDING.  
FT DOMAIN 1691 1732 EGF-LIKE 28, CALCIUM-BINDING.  
FT REPEAT 1733 1806 TGFBR 5.  
FT DOMAIN 1807 1848 EGF-LIKE 29, CALCIUM-BINDING.  
FT DOMAIN 1849 1890 EGF-LIKE 30, CALCIUM-BINDING.  
FT DOMAIN 1891 1932 EGF-LIKE 31, CALCIUM-BINDING.  
FT DOMAIN 1933 1971 EGF-LIKE 32, CALCIUM-BINDING.  
FT DOMAIN 1972 2014 EGF-LIKE 33, CALCIUM-BINDING.  
FT DOMAIN 2015 2054 EGF-LIKE 34, CALCIUM-BINDING.  
FT DOMAIN 2055 2096 EGF-LIKE 35, CALCIUM-BINDING.  
FT REPEAT 2097 2169 TGFBR 6.  
FT DOMAIN 2170 2211 EGF-LIKE 36, CALCIUM-BINDING.  
FT DOMAIN 2212 2251 EGF-LIKE 37, CALCIUM-BINDING.  
FT DOMAIN 2252 2292 EGF-LIKE 38, CALCIUM-BINDING.  
FT DOMAIN 2293 2336 EGF-LIKE 39, CALCIUM-BINDING.  
FT DOMAIN 2337 2378 EGF-LIKE 40, CALCIUM-BINDING.  
FT REPEAT 2379 2447 TGFBR 7.  
FT DOMAIN 2448 2489 EGF-LIKE 41, CALCIUM-BINDING.  
FT DOMAIN 2490 2530 EGF-LIKE 42, CALCIUM-BINDING.  
FT DOMAIN 2531 2569 EGF-LIKE 43, CALCIUM-BINDING.  
FT DOMAIN 2570 2612 EGF-LIKE 44, CALCIUM-BINDING.  
FT DOMAIN 2613 2652 EGF-LIKE 45, CALCIUM-BINDING.  
FT DOMAIN 2653 2693 EGF-LIKE 46, CALCIUM-BINDING.  
FT DOMAIN 2733 2733 EGF-LIKE 47, CALCIUM-BINDING.  
FT DISULFID 115 124 BY SIMILARITY.  
FT DISULFID 119 130 BY SIMILARITY.  
FT DISULFID 132 141 BY SIMILARITY.  
FT DISULFID 143 159 BY SIMILARITY.  
FT DISULFID 153 164 BY SIMILARITY.  
FT DISULFID 166 175 BY SIMILARITY.  
FT DISULFID 180 190 BY SIMILARITY.  
FT DISULFID 184 195 BY SIMILARITY.  
FT DISULFID 197 206 BY SIMILARITY.  
FT DISULFID 279 291 BY SIMILARITY.  
FT DISULFID 286 300 BY SIMILARITY.  
FT DISULFID 302 315 BY SIMILARITY.  
FT DISULFID 321 333 BY SIMILARITY.  
FT DISULFID 328 342 BY SIMILARITY.  
FT DISULFID 344 357 BY SIMILARITY.  
FT DISULFID 497 509 BY SIMILARITY.  
FT DISULFID 504 518 BY SIMILARITY.  
FT DISULFID 520 532 BY SIMILARITY.  
FT DISULFID 538 548 BY SIMILARITY.  
FT DISULFID 543 557 BY SIMILARITY.  
FT DISULFID 559 572 BY SIMILARITY.  
FT DISULFID 578 590 BY SIMILARITY.  
FT DISULFID 585 599 BY SIMILARITY.  
FT DISULFID 601 614 BY SIMILARITY.  
FT DISULFID 620 631 BY SIMILARITY.  
FT DISULFID 626 640 BY SIMILARITY.  
FT DISULFID 642 655 BY SIMILARITY.  
FT DISULFID 661 672 BY SIMILARITY.  
FT DISULFID 667 681 BY SIMILARITY.  
FT DISULFID 683 696 BY SIMILARITY.  
FT DISULFID 771 783 BY SIMILARITY.  
FT DISULFID 778 792 BY SIMILARITY.  
FT DISULFID 794 807 BY SIMILARITY.  
FT DISULFID 813 825 BY SIMILARITY.  
FT DISULFID 820 834 BY SIMILARITY.  
FT DISULFID 836 849 BY SIMILARITY.  
FT DISULFID 855 865 BY SIMILARITY.



Query Match 6.3%, Score 496.5; DB 1; Length 2911;  
 Best Local Similarity 23.1%; Pred. No. 9.6e-27;  
 Matches 290; Conservative 127; Mismatches 400; Indels 439; Gaps 85;

QY 190 QCVRIIRACWLASLANAKAPG---EVKMCALGTASVWVGWGTGT--CQCGLGFG 244  
 DB 508 KCIPTVSSYRC-ECNMGYK-ODANGDCIDVDECTSNPCTNGDCVNTPGSYGCKAAGFQR 565  
 QY 245 T---ACETCTEGKYGIHQDQ-ACSCVHRCSCQPLGDGS--CDDCVGWRGKCMDEIT 297  
 DB 566 TPTKACACIDIDE-----CIQNGYCKNRCVN---SDSFCICNAGF-----ELTT 609  
 QY 298 D-NC--NGTCHTSANCL---LDPDGKASCKCAAGF--RGNGVCTAINAGET---SN 343  
 DB 610 DGNKNCVDHDECTTNNCMGMCINEDGSEKICKPGFVLAIPNGRYCTVDDECQTGICMN 669  
 QY 344 GGC-----STADCKRTTPG-----NRVCV----- 363  
 DB 670 GHCINSEGSFRDC--PGLAVGMDGRVCDTHMRSTCYGIGKGVCPREPAVTNSE 726  
 QY 364 ----- 363  
 DB 727 CCCANPDYGFGRPCQPCPAKNSAEFHGLSSGVGTTVDGRDINECALDPDICANGICENL 786  
 QY 364 -----CKAGTGD--GIVLEINPCLEBNHGCDRNAECTQTGPNQAVCNCLPKYT--G 412  
 DB 787 RGSYRCNCSNGSEYEPASGRNCIDIECLVNRLCD--NGICRNT--PGSYCTCPGPFVFT 844  
 QY 413 DGKVSILNVLTN---NGGCSFPAFCVNTBODRLCTCKP--DVTGDSIVRGSIYGL 467  
 DB 845 ETECTEDINECSNCPVNGACR-----NNGSFNCECSPGSKLSSTGLTICIDSLKTC 897  
 QY 468 PKNPSTQYFFOLOEHAVERELAGPFPVTFAPLSSSFNHEPRIKMDQOGLMSQVLYRHV 527  
 DB 898 WLNIDDSCEVAVINATILKS-----ECCATLGAA-----WGP----- 930  
 QY 528 VGGQQLLDNLKVTTSATTLGEPVSISSQDVFINNKAYL-----SSDIISTNGVIH 582  
 DB 931 --CERCEGLTL--DGTGRVCLD--IRMEQCYLKMV--EDBCIHPVPGKFRMDACC 1025  
 QY 583 VIDKLSPKNNLITPKDALGRVLONLTYAANHGTYKPSKLLQDSGLSV-----ITDSIH 638  
 DB 979 ----CECEPGLTL--DGTGRVCLD--IRMEQCYLKMV--EDBCIHPVPGKFRMDACC 1025  
 QY 639 TPVTYFWPJD-----KALEALPPEQODFLFNODMNKDKLSYLFHVIDSKALASD 689  
 DB 1026 CAVGANMOTEECECPKPGTKETVETLCPRGAGF--ANRGDVLTLGRPFYDINECAFPDM 1082  
 QY 660 LPRASMKTLQGSLSVRCGTG-----SDIGELFLNDOMCRFTHRGILFDVGAV 739  
 DB 1083 CTYGKCRMTI--GSFKRCNCSGAFALDMBERNCTDIECISPDLC--GSGCIVMTPGSF 1137  
 QY 740 GIDC-----LTM-----NPTL--GGRC-DFTTFDILPGEGCSILFPKCL 777  
 DB 1138 BCECEGYESGMMKNCMDIDGCRNPLLCRGTCVNTGESF--QC-----DCPL 1186

QY 778 KSKPKGVKKKCI-YNPLPRRNVGQON-LCTVVIOTPRC-CH-GY-FMPDCQACP---- 828  
 DB 1187 GHELSRREDYCDVIDNECSISDLN--CRNGKCVNMIGTYGCSGNPGVQATPRDQGTDIE 1244  
 QY 829 -----GGPDTPCNNRMCMCDLYTPMGQCLCHTFG-----NGTAC--ELCWHGRFGDCOP 876  
 DB 1245 CMIMNGCTQCTN--SEGSY----ECSSEGYALMPGRSCADIDEC-----ENNP 1290  
 QY 877 RSCSEHGOCDEGTSGSELCETGWTAA-----SCD--TTPAVAVCTPACSVHA 924  
 DB 1291 DIC-DGGQC-TNIPGEYRLCYDGFMASMDKTCIDVNECDLNSNICMGEC----- 1340  
 QY 925 TCTENN--TCVCNLINE-GDGIT--CTVVDPCQONNGSCAKVAKSCQKTGVSCKKG 978  
 DB 1341 ---ENMKSGFICHCQGLGYVKKGTGTCTDYDECEIGAHNCMDHNSGLNIPGSKCRSG 1397  
 QY 979 YKGDGSCIEIETPCADGVNGGCHENATCMTGPKAKHCKCKSHYGDGVDCPEBOLPLDR 1038  
 DB 1398 WINGNKICIDLECSNGTH-QCSINACVNT--PGSYRCASGEGFTGDFTCSS--DYDE 1451  
 QY 1039 C-----LDNNGC-----HPDASCADL--YRDTTY-----GV 1064  
 DB 1452 CAENINLCENGGCLNVPAYKCECEMGFTTPASDSRSCQDIDECSPONICVSGTCNNLPDM 1511  
 QY 1065 FLURSPLGQYKLTDFKAKEACA--KEATTIATYNOLSYAOKAKYHICSAAGLESGRVAY 1121  
 DB 1512 FHICIDDDG-YEL--DRTGNCMDIDECADPINCYNCLVNTGRTY-CMC----- 1557  
 QY 1122 PTTVASQKCGANVGIVDYGSRAKNSKEMMDVFCYRMRKDYNTCTKAGYVDG--FSCS 1176  
 DB 1558 PDPFOLNPTG--VGCVD--NRVG-----NCYLKFGPRDGSISCN 1593

## RESULT 5

NOTC\_XENLA STANDARD: PRT: 2524 AA.  
 ID NOTC\_XENLA  
 AC P21783;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Neurogenic locus notch protein homolog precursor (XOTCH protein).  
 GN XOTCH.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90385285; PubMed=2402639;  
 RA Coffman C., Harris W., Kintner C.;  
 RT "Xotch, the Xenopus homolog of Drosophila notch.";  
 RL Science 249:1438-1441(1990).  
 RN [2]  
 RP REVISIONS TO 1759-1782.  
 RA Kintner C.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.  
 CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, M33874; AAB02039.1; -.







```

FT DOMAIN 139 176 EGF-LIKE 3.
FT DOMAIN 177 215 EGF-LIKE 4.
FT DOMAIN 217 253 EGF-LIKE 5.
FT DOMAIN 255 291 EGF-LIKE 6.
FT DOMAIN 293 329 EGF-LIKE 7.
FT DOMAIN 331 370 EGF-LIKE 8.
FT DOMAIN 372 408 EGF-LIKE 9.
FT DOMAIN 409 447 EGF-LIKE 10.
FT DOMAIN 449 486 EGF-LIKE 11.
FT DOMAIN 488 524 EGF-LIKE 12.
FT DOMAIN 526 562 EGF-LIKE 13.
FT DOMAIN 564 600 EGF-LIKE 14.
FT DOMAIN 602 637 EGF-LIKE 15.
FT DOMAIN 639 675 EGF-LIKE 16.
FT DOMAIN 677 713 EGF-LIKE 17.
FT DOMAIN 715 751 EGF-LIKE 18.
FT DOMAIN 753 789 EGF-LIKE 19.
FT DOMAIN 829 865 EGF-LIKE 20.
FT DOMAIN 867 905 EGF-LIKE 21.
FT DOMAIN 907 944 EGF-LIKE 22.
FT DOMAIN 946 982 EGF-LIKE 23.
FT DOMAIN 984 1020 EGF-LIKE 24.
FT DOMAIN 1022 1058 EGF-LIKE 25.

```

Query Match 6.2% Score 484; DB 1; Length 2703;

Best Local Similarity 23.1% Pred. No. 6.7e-26; Mismatches 382; Indels 414; Gaps 75;

```

218 MCALCTASVMDVNGTGTCCGGLGFNGTACE-----TCTEGKXGI--HCDQACS--- 264
359 IGVNMGWAGL-DCSNMTDCKQAACFGATCIDGVGSFYQCTGKRGKGLGLCHLDDATSNP 417
265 C-VHGRCSGGL-IGGSCDCDVGWRGVKCDMETTDNCLN--GTCHTSAACLLDPPDGKASC 320
418 CHADAICDTPSPINGSVACSGATGVKGVDCSEDI--DECQGGSPCEINGICVMP--GSYRC 474
321 KCAAGFRNGTVC-TAINNCGTSGNGCSTKADCKRTTPGNRVCKAGYTG-----DGIV 374
475 MCSQGF--TGPRCFETNINECESH--PCONEGSC-LDDPPTFRVCVMPGFTGTQCEIIDE 529
375 CLEINPCLENHGGC-DR-----NACCTGTGP--NOAVCN----- 405
530 C-QSNPCL-NDGTCHDKIKGFKSCALGFTGARCOINIDCQSQPNRNGICHDSIAGVS 587
406 --CLPKRYTGDGVCSL-INVCLTN--NGGCSPPAFECNTBEDQRICTCKPDYTGDIYC 459
588 CECPPGYTGT--CEININDCDSNPCRKGC-----IDVNSFKCLCDPGYT--GYIC 636
460 RGSITGELPKNP-----STQYFPOLQEHAVRELAGPGFTYFAFALSSFFNHEPR 510
637 OKOI-NECESNPCQFDGHCQDRVSGTYCCQ-----AG-----TSCKNEEVN 678
511 KMDQDGLMSQVLRHVHVCQOQLLDNLKVTTSATTLQCEPVSISVSQDTVFNNKAKYL 570
679 NE-----CHSNPC-----NNCAT--- 691
571 SSDIISTNGVHIVDKLLSPKNLLI--TPKDALGRLVIONLTTVAANHGYTFESKLIDS 627
692 -----CIDGINSYKQCQVGFGTGQHEKKNVDECISPCANNG----- 728
628 GLLSVITDSIHTPVYFWMTDKALEALPREQODFLFNQONKDKLKYLFHYIRDSKALA 687
729 ---VCIDOV-----NGKRCCEPREFYVAHCHLSDVDECA 758
688 SLPSPASAKTIQG-SELSVRCGTG-----SDIGELFLNEQMCFFIHRGLFLDVGAV 739
759 SN-PCVNERCDEGIDNEFICHCPGYTGKRCLEIDECSSNP--CQ--HGCTCYDKLNAF 813
740 GIDCLAMPNTLGLRCDTFTTFDIPGEC--GSCI-----PTPK-CPLAKSKR 782
814 SCQC--MPCYTQKCKETNIDCVTNPCGNGGTCIDKVNQYKCVKQVFPFGRCESKMDPC 871

```

```

QY 783 GYKK-----KCI--YNPLFR-----RVNEGCONL----- 805
DB 872 ASNRCKNEAKCTPSSNFLEDFSCCTCKLGYRCYDIEDIDECSSLPKRNASCILNAGSVR 931
QY 806 GTVVIOTPRCHGVEYEMPC-----QACPGGPDTPCNRNMGKCRDLYTPMG--QCCLHTGFG 859
DB 932 CL-----CTKGYEGRDCAIMTDDC--ASFPQCGGTCLD---GIGDYSCLCVGDFGG 978
QY 860 TACEL-----CHNGRF-----GPDQPRGSEHGQD 886
DB 979 KHCIEDINECLSQPCQNGATCSQYVNSYCTGCLPFGSGINCOITNDEDTRESSCLNGSCSI 1038
QY 887 EGTGSGEFLCTGWTAAACDPTPAFAVC-TPACSVHATCTENN--TCVCNLNANEGDG 942
DB 1039 DGINCY-NCSCLAGSYGANCQYK---LNKCDNSPLCLNGATCHRONNRYTHCHPSGTTGR- 1093
QY 943 ITCT-VVDFCKQNNNGCAVAVKCSQKGTQVSSCKKGYKGDYSC-IEIDPCADGV---- 996
DB 1094 -QSEYVDMCGQS--PCENGATCSQMKHQFSCKCSAGMTGK--LCDVQYITSCODADARRKG 1148
QY 997 -----NGGCHENATCMTGKRGKRCCKSHYVDGVDCERPOLPLDRCLDNGOCHP 1048
DB 1149 LSLRLQCNNGTCKDYG-----NSHVYCSQGYAGS--YCKQK--IDEC--QSQPCQN 1194
QY 1049 DASCADLYFDQITVGVFHLRSLPGYKLFDPKAKECAKEAATATYNDLSYAQRKAYHL 1108
DB 1195 GGTGRDL-----IGAYEDQCRGFGQNCLELIDCAFPN----- 1229
QY 1109 CSAGWLESGRVAVPTTYASQKCGANVGTVDYGRANKSEMDVFCY-----RMKDVN 1161
DB 1230 CQNGGTCHDRV-----MNFSCSPGPTGMII--CEINKDKRKGACHNNNGSCIDRVGPE 1282
QY 1162 CTCKAGYVGDGFSGSLNLOVL 1183
DB 1283 CYCQPGFV--GARCEGINECL 1302

```

RESULT 7

ID	FBN1_PIG	STANDARD	PRT	2871 AA.
AC	Q9TV36;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Fibrillin 1 precursor.			
GN	FBN1.			
OS	Sus scrofa (pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lung;			
RX	MEDLINE=99156858; PubMed=10036187;			
RA	Bieri N.J., Eldadah Z.A., Moore C.S., Stetten G., Spencer F.,			
RA	Dietz H.C.;			
RT	Revised genomic organization of FBN1 and significance for regulated			
RT	gene expression.";			
RL	Genomics 56:70-77(1999).			
CC	- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS			
CC	THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE			
CC	LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.			
CC	- PFM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER			
CC	FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE			
CC	MICROFIBRILS.			
CC	- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING			
CC	EGF-LIKE DOMAINS.			
CC	- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.			
CC	CC This swiss-prot entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF073800; AAD50328.1; -  
DR HSSP; P35555; IAPJ.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR001438; EGF\_II.  
DR InterPro; IPR002212; TB.  
DR Pfam; PF00008; EGF; 45.  
DR Pfam; PF00683; TB; 9.  
DR PRINTS; PR00010; EGFBL00D.  
DR SMART; SM00179; EGF\_CA; 40.  
DR SMART; SM00001; EGF\_Like; 6.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_2.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_2.  
DR PROSITE; PS01186; EGF\_2; 36.  
DR PROSITE; PS01187; EGF\_CA; 41.  
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;  
Repeat; Signal; Multigene family.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 2871 FIBRILLIN 1.  
FT DOMAIN 81 112 EGF-LIKE 1, NON-CALCIUM BINDING.  
FT DOMAIN 115 146 EGF-LIKE 2, NON-CALCIUM BINDING.  
FT DOMAIN 147 178 EGF-LIKE 3, NON-CALCIUM BINDING.  
FT DOMAIN 246 287 EGF-LIKE 4, CALCIUM-BINDING.  
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM-BINDING.  
FT REPEAT 330 390 TGFBR 1.  
FT DOMAIN 392 446 PRO-RICH.  
FT DOMAIN 449 489 EGF-LIKE 6, NON-CALCIUM BINDING.  
FT DOMAIN 490 529 EGF-LIKE 7, CALCIUM-BINDING.  
FT DOMAIN 530 571 EGF-LIKE 8, CALCIUM-BINDING.  
FT DOMAIN 572 612 EGF-LIKE 9, CALCIUM-BINDING.  
FT DOMAIN 613 653 EGF-LIKE 10, CALCIUM-BINDING.  
FT REPEAT 654 722 TGFBR 2.  
FT DOMAIN 723 764 EGF-LIKE 11, CALCIUM-BINDING.  
FT DOMAIN 765 806 EGF-LIKE 12, CALCIUM-BINDING.  
FT DOMAIN 807 846 EGF-LIKE 13, CALCIUM-BINDING.  
FT DOMAIN 910 951 EGF-LIKE 14, CALCIUM-BINDING.  
FT REPEAT 952 1027 TGFBR 3.  
FT DOMAIN 1028 1069 EGF-LIKE 15, CALCIUM-BINDING.  
FT DOMAIN 1070 1112 EGF-LIKE 16, CALCIUM-BINDING.  
FT DOMAIN 1113 1154 EGF-LIKE 17, CALCIUM-BINDING.  
FT DOMAIN 1155 1196 EGF-LIKE 18, CALCIUM-BINDING.  
FT DOMAIN 1197 1237 EGF-LIKE 19, CALCIUM-BINDING.  
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCIUM-BINDING.  
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCIUM-BINDING.  
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCIUM-BINDING.  
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCIUM-BINDING.  
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCIUM-BINDING.  
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCIUM-BINDING.  
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCIUM-BINDING.  
FT REPEAT 1528 1605 TGFBR 4.  
FT DOMAIN 1606 1647 EGF-LIKE 27, CALCIUM-BINDING.  
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCIUM-BINDING.  
FT REPEAT 1689 1765 TGFBR 5.  
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCIUM-BINDING.  
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCIUM-BINDING.  
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCIUM-BINDING.  
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCIUM-BINDING.  
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCIUM-BINDING.  
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCIUM-BINDING.  
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCIUM-BINDING.  
FT REPEAT 2055 2126 TGFBR 6.  
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCIUM-BINDING.  
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCIUM-BINDING.  
FT DOMAIN 2206 2246 EGF-LIKE 38, CALCIUM-BINDING.  
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.  
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.  
FT REPEAT 2333 2401 TGFBR 7.  
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.

FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.  
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.  
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.  
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.  
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.  
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.  
FT DISULFID 89 94 BY SIMILARITY.  
FT DISULFID 89 100 BY SIMILARITY.  
FT DISULFID 102 111 BY SIMILARITY.  
FT DISULFID 119 129 BY SIMILARITY.  
FT DISULFID 123 134 BY SIMILARITY.  
FT DISULFID 136 145 BY SIMILARITY.  
FT DISULFID 150 160 BY SIMILARITY.  
FT DISULFID 154 166 BY SIMILARITY.  
FT DISULFID 168 177 BY SIMILARITY.  
FT DISULFID 250 262 BY SIMILARITY.  
FT DISULFID 257 271 BY SIMILARITY.  
FT DISULFID 273 286 BY SIMILARITY.  
FT DISULFID 292 304 BY SIMILARITY.  
FT DISULFID 299 313 BY SIMILARITY.  
FT DISULFID 315 328 BY SIMILARITY.  
FT DISULFID 453 465 BY SIMILARITY.  
FT DISULFID 460 474 BY SIMILARITY.  
FT DISULFID 476 488 BY SIMILARITY.  
FT DISULFID 494 504 BY SIMILARITY.  
FT DISULFID 499 513 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 534 546 BY SIMILARITY.  
FT DISULFID 541 555 BY SIMILARITY.  
FT DISULFID 557 570 BY SIMILARITY.  
FT DISULFID 576 587 BY SIMILARITY.  
FT DISULFID 582 596 BY SIMILARITY.  
FT DISULFID 598 611 BY SIMILARITY.  
FT DISULFID 617 628 BY SIMILARITY.  
FT DISULFID 623 637 BY SIMILARITY.  
FT DISULFID 639 652 BY SIMILARITY.  
FT DISULFID 727 739 BY SIMILARITY.  
FT DISULFID 748 748 BY SIMILARITY.  
FT DISULFID 750 763 BY SIMILARITY.  
FT DISULFID 769 781 BY SIMILARITY.  
FT DISULFID 776 790 BY SIMILARITY.  
FT DISULFID 792 805 BY SIMILARITY.  
FT DISULFID 811 821 BY SIMILARITY.  
FT DISULFID 816 830 BY SIMILARITY.  
FT DISULFID 832 845 BY SIMILARITY.  
FT DISULFID 946 926 BY SIMILARITY.  
FT DISULFID 921 935 BY SIMILARITY.  
FT DISULFID 937 950 BY SIMILARITY.  
FT DISULFID 1032 1044 BY SIMILARITY.  
FT DISULFID 1039 1053 BY SIMILARITY.  
FT DISULFID 1055 1068 BY SIMILARITY.  
FT DISULFID 1074 1086 BY SIMILARITY.  
FT DISULFID 1081 1095 BY SIMILARITY.  
FT DISULFID 1097 1111 BY SIMILARITY.  
FT DISULFID 1117 1129 BY SIMILARITY.  
FT DISULFID 1124 1138 BY SIMILARITY.  
FT DISULFID 1140 1153 BY SIMILARITY.  
FT DISULFID 1159 1171 BY SIMILARITY.  
FT DISULFID 1201 1212 BY SIMILARITY.  
FT DISULFID 1208 1221 BY SIMILARITY.  
FT DISULFID 1223 1236 BY SIMILARITY.  
FT DISULFID 1242 1254 BY SIMILARITY.  
FT DISULFID 1249 1263 BY SIMILARITY.  
FT DISULFID 1265 1278 BY SIMILARITY.  
FT DISULFID 1284 1296 BY SIMILARITY.  
FT DISULFID 1291 1305 BY SIMILARITY.  
FT DISULFID 1307 1320 BY SIMILARITY.  
FT DISULFID 1326 1339 BY SIMILARITY.  
FT DISULFID 1333 1348 BY SIMILARITY.  
FT DISULFID 1350 1361 BY SIMILARITY.  
FT DISULFID 1367 1380 BY SIMILARITY.  
FT DISULFID 1374 1389 BY SIMILARITY.  
FT DISULFID 1391 1402 BY SIMILARITY.

```

FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.
FT DISULFID 1674 1687 BY SIMILARITY.
FT DISULFID 1770 1782 BY SIMILARITY.
FT DISULFID 1777 1791 BY SIMILARITY.
FT DISULFID 1793 1806 BY SIMILARITY.
FT DISULFID 1812 1824 BY SIMILARITY.
FT DISULFID 1818 1833 BY SIMILARITY.
FT DISULFID 1835 1847 BY SIMILARITY.

```

Query Match 6.28: Score 483.5; DB 1; Length 2871;  
 Best Local Similarity 22.38; Pred. No. 7.9e-26;  
 Matches 259; Conservative 116; Mismatches 342; Indels 443; Gaps 69;

```

QY 189 POCVRIITRACWL-ASLAHNAK--PAPG-----EYKMCALGTASV 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 PRVLPVAVTDCQLFRILCHNGRCIPFGSYRCENKGFOLDLAGECIDVECEKNPCAG 501
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 227 WDVYNGTG--TCQGLGFNGTACECTEGKYGIHCDQ-ACSCVHGRCQGFADGS--CD 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 502 GECINNGSYTCQCRPGYSTL--TRTECRDIDELQNGRICNNGCIN--TDGSPFCV 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 CDVQMR---CVKQDMETTDNC-----NGTCHTSANCLLDPDGKASCAAGFR--G 328
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 557 CNAGFHVTRDKNCE---DMDECSIRNMCLNGMC-----INBDGSEKCIKPKFOLAS 606
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 329 NGVTCTAINACETS-----NGGC-----KRTPTGNRV-- 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 DGRYCKDINEETSIGICMNGRCVNTDGSYRCFCPLGAVLGDGRVCDTHMRSTCYGYK 666
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 -----STKADC-----KRTPTGNRV-- 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 667 RGQCVKPLFGAVTKSECCASTEYAFGEBCOPCPSONSAEYQALCSSGCPMTSAGSDINE 726
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 362 -----CVKQAGYTG--GIVCLEINPCLENGCCDRAEACTQTG 398
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 727 CALDPDICPNICENLRGTAKYCICNSGEVDSGTCKNCVDINECVLNSLLCD--NGQCRNT- 784
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 399 PNOAVCMCLPK---YTDGKVCSLINVCLTN---NGGC--SPAFPCNTEDDQRICTCKP 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 785 PGRSVCTC-FKGFITYKPKDLKTCEDIDECSSPCINGVCKNSPGSF-----ICQSS 834
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 451 DYTGD--GIYCGRSTYV-----ELPKNPSTSOYFPQLOEHAVERLAGP-----G 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 835 ESTIDPKTICIEIKICMOTIIDGRCEININCAT-----LKSQCCSSLGAAWGPST 888
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 493 PTYFALASSFNHRIKMDQGLMSOYLRIHVVGCOQLLDNLKVTTSATTLQGEPEV 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 889 PCQVDPICGKGY--RIR-----CTQCEDIDECCEFP----- 918
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 553 SISVQDTPFINNEAKYLSSDIISTNGVIVHIDLSPKNLLIPKDALGVNLNTTVA 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 919 --GVCKNGKLCVNSKGF-----KQCPNGMNTL--DATGICIDIRL-- 955
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 613 ANHGTYFESKLIQDSGLLSV---TDSIHTPVTPWPTDKALALP--PEQDFL-- 662
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 956 ----ETCFLEKDEDECTLPVVGGRHMDACCCSVGAAMGTECECPCPRNPREYELCPRG 1011
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 663 --FMQDKKDKLSLKFHVIRDSKALASDLPRSSMKTLLQCSSELSVRCGTSDIGELFLN 720
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1012 PGFAKTEITMGKRP--FKDINECKMIPNLTCHGKCRMTI--GSPKCRCDSGFALDS--E 1064
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 721 EOMCRFTIHRGLLPDVGAAGIDCLLMNPTLGGRCDTFTTEDIPECGSCITFP-----KC 775
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1065 ERNC-----IDIDECRISPDLGCR-----GQCVMTPPDGECK 1097
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 776 PLKSKPKG--VKKCIYNPLPFRNRBGCN---LCT--VVIQTP-----RCHGEYF-P 822
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1098 D-EGYESGFMMKNCNM-----DIDECQDPRLCIRGCVLNTGEGSYRCFCPSGHQMSF 1148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 823 DCOAC-----PGSPDTPCANNRGRCRLYPMQGL 852
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1149 NISACIDINECELSAHLCPHGRCVNLGKYORARNPGYHSP--DRLECYDI--DECS 1202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 853 CHTFNGFACELCWHGRCPCOP-----RSCSEHGQDE-----GIRSG 893
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1203 IMNGGCEFFCINS-EGSYECQOPGFALMPQORCTDIDEDPNICDGGQCTNIEGEY 1251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 894 ECLCETGNTAA-----SCDPTAVFAVCTPACSVHATCTENN---TCVCNINYE 939
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1262 RCLCYDGFMASEDKMKTCVDVNECD-----LNPNICLSCTC-ENRKGSFICHQDMGYS 1312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 940 G-DGIT-CTVVDPFCKONNGGCAKAKCSQKGTQVSCSKKGYKDDGYSCIEIDPCADGVN 997
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1313 GKRGKTGTDINECEIGAHCNDRAHVCTNTAGSFNCSPGMIQDGIKCTDLDECSTMGTH 1372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 998 GGCHEHATCRMTGPKHKRCEKSHYVGVCYDCEPEQLPLDCLQD-----NGQCHPADSC 1052
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1373 -MCSQHADCKNT-MGSTRCLCKEGYTGDFTC---ADLDECSENVKLCGNVOC----- 1420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1053 ADLYFQDTTVGVFHLRSPUG 1072
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1421 --LV-----APCGYHCEYDMG 1434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 8
NTCL_HUMAN STANDARD; PRT: 2444 AA.
AC P46531;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch protein homolog 1 precursor (Translocation-
DE associated notch protein TAN-1) (Fragment).
GN NOTCH1 OR TAN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91347367; PubMed=1831692;
RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA Smith S.D., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms.";
RL Cell 66:649-661(1991).
CC -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION IN
CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC IN SOME T-CELL NEOPLASMS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC IS FOUND MAINLY IN LYMPHOID TISSUES.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/)

```





Query Match 6.1%; Score 482.5; DB 1; Length 2444;  
 Best Local Similarity 21.7%; Pred. No. 7.5e-26;  
 Matches 269; Conservatve 120; Mismatches 445; Indels 407; Gaps 75;

127 EKVLEIKNKNNDNDITIVGEC--GKCSQAPRPLE-----TKPLRETR----- 169  
 Db ENIDDCGNCKNG-----GACVDGVNTYNPCPPEMTGQYCTEDVDDECOLMNAQNG 309  
 170 -----KCIYSIFMGKRSVFIGCQPCVRTI---ITRACMLASLHNAK----- 210  
 Db 310 GTCNHTHGNGVNCVNWMTGE-----DCSENIDDCASACFHCATGCHDRASFYCE 360  
 211 -PABGEVKKALGTASY-----WDGVNGTGTCCGLGFNGTAC---ETCTEGK 254  
 Db 361 CPHGRTGLLCLINDACISNPCNEGSCDTPNVNGKALICTCPSTGYTPACSDVDDECSLG 419  
 255 YGHCADACSCVHRCGSGPLGDSGCCDGVGMRGVKCDMETTONG-NGTHTSANCILD 313  
 Db 420 -ANPCEHAGKCN-----TLGSFECCLGTYGTPRC--EIDVNECVSNPCQNDATC-LD 469  
 314 PDGKASCKCAAGFRNGTVCATAINA--CETS---NGCSTKADCKRTTPGNRVCCVCKAG 367  
 Db 470 QIGFQCMCMKPGYRG---VHCEVNTDECASSPCLHNGRCLDKIN-----EPCECPYTG 519  
 368 YTGDDIVC-LEINPCLNHHGCGDRNAECTQTGPNAVAVCNCLPKYTGDKYCSLINVCLTN 426  
 Db 520 FTGH--LCQYDVDEACAST--PCKNGAKCLD-GPNTYTCVCTEGYTG-----THCEVD 566  
 427 NGGSPAPFCNYTPQDRI---CTCKPDYTGDIIVRGSLTYGLPKNPSTSYRFPLOE 482  
 Db 567 IDECDPPD-CHYGSCKDGVATFTCLCRPGYTGHH--CETNI-NECSSQPCRLRGTCODPD 622  
 483 HA-----VRELAGRPPTVFAPLSSSFNHEPRIKDMWDQGLMSQVLRYHVYVCCQLLDN 537  
 Db 623 NAYLCFCLKTGTG-----NCEINLND-----CASSPCDS 652  
 538 LKVTATSTLQGEFVSISVSODIVFINNEAKVLSDDITSTNGVINVHIDKLSPPNLLTP 597  
 Db 653 ---GTCLDKIDGYECACEPEGYTGSMSNSNIDECAGNPNCHNG-----TC 693  
 598 KDALGRVLONTTVAANHGYTKFKSKLLJODSGLSVITDSIHTPYTVFWPTDKALEALPPE 657  
 Db 694 EDGJ---NGFTCRPEGY-----HDPYCLSEVNECNSNPCV----- 726  
 658 QODFLFNQDNKDKLSYLFKHV-----IRDSKALSDLPKRSASWKTLOGSELSVR 707  
 Db 727 -----HGACRDSLNGY-KCCDDPCWMSGTNDINNNECESPCVNGGCKMT--SGIVCT 778  
 708 CGTG-----SDIGELFLNQMCRFIHRGLFDVGAAYGIDCLLMMFTLGRCDTFTTF 760  
 Db 779 CREFFSGPNQCTININECASNP--C--LNKGTICIDVAGYKCNCLL--PYTGATCEVVLAP 832  
 761 DIPBEC---GSC-----IFTPKCPILSKRKGVK-----KKCIYNPLPFRRNWEGC 802  
 Db 833 CAPSPCRNGGCRGROSEDEYSFSCVCP---TAGAKGOTCEVDINECVLSPC--RHGASC 885

803 QNLCTVVIQTPRCCHGYEMPDQACPGGP-----TPCNRNGKCRD-LYTPMGOC 851  
 Db 886 QN-----THGXRYCHQAGYSGRNCEITDIDCRPRPCHNGSGCTGINTAF--C 932  
 852 LCHTGFNGTAGE-----LCMHGRFG-----PDCOPRS 878  
 Db 933 DCLGFGRFTFCEEDINECASDPCKRNKANCITDCVDSYTCCTCPAGFSGHICENNFTDCTESS 992  
 879 CSEHGQCEDEGTTGSGECLCTGWTAAASCDTPTAFAANC--TPACSVHATCTEN---NVCVC 934  
 Db 993 CFNGGTCVDGI-NSFTCLCPRGFTGYSYQ---HVVNECDSPRLGLGTCODGRGLHRCCTC 1048  
 935 NLNNEGDDITC-TVVDECKONNGCAKAVKCSQKGTQVSCSCCKRGYK-----DGYSCIE 988  
 Db 1049 PQGTGTP--NQNILVHNC--DSSPCKNGKCMQHTHTQYRCBCPSGWTGLYCDVSVSC-E 1103  
 989 IDPCADGVNNG--CHEHATCMRTGPKHKHCKCKSHVVG---DQVD-CEPBQLPLDRCLQ 1041  
 Db 1104 VAAOROGVDVARLCOHGLICVDAG-NTHHRCQAGYTGSGYCEDLVDECSF----- 1152  
 1042 DNGCHPDASCADLYFDQTTGVFHLKSLPGQYKLTFDKAKEACAKKATATATYNQLSYA 1101  
 Db 1153 --SPQNGATCTDY-----LGGYSCKCVAGYHGVNCSSEIDECLS----- 1190  
 1102 QKARYHLCSAGMLESGRVAYPTTYASQKCGANVGI-----VDYGRANKSEM 1149  
 Db 1191 -----HPCQNG---CTCLDLPLNTYK--SCPRGTQGVHCEINVDONCPVDPVSNKSPK--- 1238  
 1150 WDVFCY-----RMKDVNCTCKAGYVGDGFSGSGNLQVL 1183  
 Db 1239 ---CFNNGTCVDPQGVGYSCTCPPGFVGE--RCGADVNECL 1273

RESULT 9  
 FBNI\_BOVIN  
 ID FBNI\_BOVIN STANDARD; PRT; 2871 AA.  
 AC P98133;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Fibrillin 1 precursor (MP340).  
 GN FBNI.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_Taxid=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SKIN;  
 RX MEDLINE=95137597; PubMed=7835900;  
 RA Tilstra D.J., Potter K.A., Byers P.H.;  
 RT "Sequence of the coding region of the bovine fibrillin cDNA and  
 RT localization to bovine chromosome 10.";  
 RL Genomics 23:480-485(1994).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=96132851; PubMed=8557636;  
 RA Gibson M.A., Hatzinikolas G., Kumaratillake J.S., Sandberg L.B.,  
 RA Nicholl J.K., Sutherland G.R., Cleary E.G.;  
 RT "Further characterization of proteins associated with elastic fiber  
 RT microfibrils including the molecular cloning of MAGP-2 (MP25).";  
 RJ J Biol. Chem. 271:1096-1103(1996).  
 CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
 CC -!- BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS  
 CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.  
 CC -!- FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER  
 CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE  
 CC MICROFIBRILS.  
 CC -!- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING  
 CC EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.

CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	CC	the European Bioinformatics Institute. There are no restrictions on its
CC	CC	use by non-profit institutions as long as its content is in no way
CC	CC	modified and this statement is not removed. Usage by and for commercial
CC	CC	entities requires a license agreement (See <a href="http://www.ebi.ac.uk/announcements/eblsb-sib.ch">http://www.ebi.ac.uk/announcements/eblsb-sib.ch</a> ).
CC	CC	or send an email to <a href="mailto:license@eblsb-sib.ch">license@eblsb-sib.ch</a> .
DR	EMBL; L28748; AAA74122.1; -.	
DR	HSPD; P35555; IAPJ.	
DR	InterPro: IPR000152; Asx_hydroxyl.	
DR	InterPro: IPR000561; EGF-like.	
DR	InterPro: IPR001881; EGF_Ca.	
DR	InterPro: IPR001438; EGF_II.	
DR	InterPro: IPR002212; TB.	
DR	Pfam; PF00068; EGF; 46.	
DR	Pfam; PF00683; TB; 9.	
DR	PRINTS; PR00010; EGFBLOD.	
DR	SMART; SMO0179; EGF_CA; 42.	
DR	SMART; SMO0001; EGF_like; 4.	
DR	PROSITE; PS00010; ASX_HYDROXYL; 43.	
DR	PROSITE; PS00022; EGF_1; 2.	
DR	PROSITE; PS01165; EGF_2; 38.	
DR	PROSITE; PS01187; EGF_CA; 45.	
KW	Extracellular matrix; Calcium-binding; Glycoprotein; Egf-like domain;	
KW	Repeat; Signal; Multigene family.	
FT	SIGNAL	1 27
FT	CHAIN	28 2871
FT	DOMAIN	81 112
FT	DOMAIN	115 146
FT	DOMAIN	147 178
FT	DOMAIN	246 287
FT	DOMAIN	288 329
FT	REPEAT	330 390
FT	DOMAIN	392 446
FT	DOMAIN	449 489
FT	DOMAIN	490 529
FT	DOMAIN	530 571
FT	DOMAIN	572 613
FT	DOMAIN	613 653
FT	DOMAIN	654 722
FT	REPEAT	723 764
FT	DOMAIN	765 806
FT	DOMAIN	807 846
FT	DOMAIN	910 951
FT	DOMAIN	952 1027
FT	REPEAT	1028 1069
FT	DOMAIN	1070 1112
FT	DOMAIN	1113 1154
FT	DOMAIN	1155 1196
FT	DOMAIN	1197 1237
FT	DOMAIN	1238 1279
FT	DOMAIN	1280 1321
FT	DOMAIN	1322 1362
FT	DOMAIN	1363 1403
FT	DOMAIN	1404 1445
FT	DOMAIN	1446 1486
FT	DOMAIN	1487 1527
FT	DOMAIN	1528 1605
FT	REPEAT	1606 1647
FT	DOMAIN	1648 1688
FT	DOMAIN	1689 1765
FT	REPEAT	1766 1807
FT	DOMAIN	1808 1848
FT	DOMAIN	1849 1890
FT	DOMAIN	1891 1929
FT	DOMAIN	1930 1972
FT	DOMAIN	1973 2012
FT	DOMAIN	2013 2054
FT	REPEAT	2055 2165
FT	DOMAIN	2127 2165
FT	DOMAIN	2166 2205
FT	DOMAIN	2206 2246

FT	DOMAIN	2247	2290	EGF-LIKE 39, CALCION-BINDING.
FT	DOMAIN	2291	2332	EGF-LIKE 40, CALCION-BINDING.
FT	REPEAT	2333	2401	TGFBP 7.
FT	DOMAIN	2402	2443	EGF-LIKE 41, CALCION-BINDING.
FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCION-BINDING.
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCION-BINDING.
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCION-BINDING.
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCION-BINDING.
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCION-BINDING.
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCION-BINDING.
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	257	271	BY SIMILARITY.
FT	DISULFID	273	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.
FT	DISULFID	299	313	BY SIMILARITY.
FT	DISULFID	315	328	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	460	474	BY SIMILARITY.
FT	DISULFID	476	488	BY SIMILARITY.
FT	DISULFID	494	504	BY SIMILARITY.
FT	DISULFID	499	513	BY SIMILARITY.
FT	DISULFID	515	528	BY SIMILARITY.
FT	DISULFID	534	546	BY SIMILARITY.
FT	DISULFID	541	555	BY SIMILARITY.
FT	DISULFID	557	570	BY SIMILARITY.
FT	DISULFID	576	587	BY SIMILARITY.
FT	DISULFID	582	596	BY SIMILARITY.
FT	DISULFID	598	611	BY SIMILARITY.
FT	DISULFID	617	628	BY SIMILARITY.
FT	DISULFID	623	637	BY SIMILARITY.
FT	DISULFID	639	652	BY SIMILARITY.
FT	DISULFID	727	739	BY SIMILARITY.
FT	DISULFID	734	748	BY SIMILARITY.
FT	DISULFID	750	763	BY SIMILARITY.
FT	DISULFID	769	781	BY SIMILARITY.
FT	DISULFID	776	790	BY SIMILARITY.
FT	DISULFID	792	805	BY SIMILARITY.
FT	DISULFID	811	821	BY SIMILARITY.
FT	DISULFID	816	830	BY SIMILARITY.
FT	DISULFID	832	845	BY SIMILARITY.
FT	DISULFID	914	926	BY SIMILARITY.
FT	DISULFID	921	935	BY SIMILARITY.
FT	DISULFID	937	950	BY SIMILARITY.
FT	DISULFID	1032	1044	BY SIMILARITY.
FT	DISULFID	1039	1053	BY SIMILARITY.
FT	DISULFID	1055	1068	BY SIMILARITY.
FT	DISULFID	1074	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1111	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1124	1138	BY SIMILARITY.
FT	DISULFID	1140	1153	BY SIMILARITY.
FT	DISULFID	1159	1171	BY SIMILARITY.
FT	DISULFID	1166	1180	BY SIMILARITY.
FT	DISULFID	1182	1195	BY SIMILARITY.
FT	DISULFID	1201	1212	BY SIMILARITY.
FT	DISULFID	1208	1221	BY SIMILARITY.
FT	DISULFID	1223	1236	BY SIMILARITY.
FT	DISULFID	1242	1254	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1265	1278	BY SIMILARITY.
FT	DISULFID	1284	1296	BY SIMILARITY.
FT	DISULFID	1291	1305	BY SIMILARITY.
FT	DISULFID	1307	1320	BY SIMILARITY.

Query Match 6.1% Score 481; DB 1; Length 2871;  
 Best Local Similarity 20.8%; Pred. No. 1,2e-25;  
 Matches 276; Conservative 128; Mismatches 383; Indels 542; Gaps 77;

QY 125 GLEKVLLEOKRCD-----NNDTIIVRG-----ECGKCSQOAPCPLETPPLAET 168  
 DB 1100 GYESGFMMKMKMDIDECQRPDLRCGGVCLNTEGSRCECPGHQLAP---NISACIDI 1156  
 QY 169 KCIYSIYFMKRSYFIC-OPOCVRTIITRACMLASLAHNAKPAPEGVKKMCLGASVW 227  
 DB 1157 NECELSAIL-----CPHGRCVNLIGKYQACNPGYHS---TPDLFCVVDIDECSTIM 1204  
 QY 228 DG-----VNGTGT--COGGLGF-----NGTACE--TCTE--GKYGTH 258  
 DB 1205 NGGCEFTCTNSEGSEYECSCOPGFALMPDRSCITDIDECEDNPINIDGGQCTNIPGEYRCL 1264  
 QY 259 CDQA-----CS-----CVHGRCSOGPLDGSODDOVGARGVK---C--- 291  
 DB 1265 CYDGFMASEDKTKTCDVNECLNPICLSGTC-ENTKGSFICHCMDGYSKGKGTGCTDI 1323  
 QY 292 -DMEITTDNC--NGTCHTSANCLDPPDGKASCKACAGFRGNVTGTAINACETSMGCGST 348  
 DB 1324 NECEIGAHNCORHVAVCTMTA-----GSFKGSCSPGWIGIGIKTDTLDECSNGTHMSQ 1376  
 QY 349 KADCKRTTPGKRVCCVCAKGYTGDGIVCLINPCLENHGGCDRMAECTGTGPNOAVCNC-- 406  
 DB 1377 HADCKNTM-GSYRCLCKBGTGTGDTCTDLDECSENLMICG-NGQCL-NAPGVRGCECDM 1433  
 QY 407 --LPKYGDGV---CSLINVCL-----TNNGGCSPFAFC 436  
 DB 1434 GFVP--SADGRACEDIDECSLPNICVFCTCHNLPGLFRCECEIGELDRSGMCTDVNEC 1491  
 QY 437 -----NYTEDODRITCKPDY-----TG 454  
 DB 1492 LDPTTCISGNCVNTPGST-----CDCPRPFLNPTRVGCVDPHSGACIYDIRRGNG 1545  
 QY 455 D-----GIYCR-GSIYG--ELPKPNDSTOYFPQLOEHAHVRELAFG--- 492  
 DB 1546 DTACSNIEGVSVKASCCCSLGKAMGPCELCPPVNTSEY-----KILCPGGEGF 1595  
 QY 493 ---PFTVAPLSSSFNHEPRIKMDQOGLMSOVLRYHVVGQOL--LLDNLKVTTSATTL 547  
 DB 1596 RPNPITVY-----LEDIDE-----COELPGLCQGGCGCINTFESF 1629  
 QY 548 QGE-PVSISSVQDTVF-----INNEAKVLSDIISTINVIHVID 585  
 DB 1630 QCRCPGYLLMEDIRVYCDVNECEPPTGICGPTCQNTYGCNTTICPPRYMOYVNGNCCMD 1689  
 QY 586 --KLSPKNLLITPRDALGRVQLNLTVAANHGYTKFSKLLIODSGLSVTDSHTPTVY 643  
 DB 1690 MRRSLCYRNYADNQTCGGELLFNTKMKCCSYN-----IGR 1727  
 QY 644 FWPIDKALEALPPEGGDPLFQDQNKDKLSYLKHFVIRDSKALASDLPRSASMKTLQSGE 703  
 DB 1728 AM--NKPCGCP-----IPSTDEPVLGSGO 1751

QY 704 -----LSVRGTSGDIGELFLNEQMKRFTIHRQLLPDVGAVAGIDCLMNPPLGRCRDTFT 758  
 DB 1752 RPFGEIDYITGLPVIDE-----CR----- 1771  
 QY 759 TFDIDGEC--GSCI-----PTPKPLKSPKGVKKKCIYN-PLPRRNVGCON--LCTV 808  
 DB 1772 --EIPGVCENGVCIMMVSFRCECHV-----GFYFNKDLTLCVEDIDECQNPVCO- 1819  
 QY 809 VIQTRCGH--GYFPRDOACGCPDT---PCNNRGMCRDLTPMGQCLCHTGFSTACE 863  
 DB 1820 --RNAECINTASRYCDCK--PGYRFTSTGOQDNNECOET----- 1856  
 QY 864 LCMHGRFEPDCCPRSCSEHQDCDEGITSGBELCTFGTMTAASCDPPTAVFAVCPACSVH 923  
 DB 1857 -----PNICS-HGQCIDIV--GSFYCLCHTGRKTANDQYMCIDLINCEBDAQCN 1902  
 QY 924 ATCTENNT-----CYCN-----LNYEDGITTCTVDPFCOMNGGCAKVAKCSOKTGVSCS 974  
 DB 1903 GTC--RNTIGSFNCRGNHGFILSHNND---CIDVDECATGNGLCRNGQCIINTVGSFQOQ 1957  
 QY 975 CKKGYK--GDGYSCTEI-----DP--CADGVNCGCHEHATCRMTGPKHCKRSHYVGD 1025  
 DB 1958 CMEGYEVAPDGRCTVDINECLDLPKRCAPG-----TCQNL-DGSTKICPPGYSLQ 2007  
 QY 1026 GVDCPEQLPDLRCLQDNQCHPDASCADLYFQDTPVGVFHLRSPPLGQYKLPFDKAKBAC 1085  
 DB 2008 NDKCE--DIDCEVEPEIC-ALGTCSN-----TESFKLCIDCG---FSLSTGRC 2053  
 QY 1086 AKEAATYATYNOLSYAKAKYHLCSAGMLESGRAYPTT--YASOKGANYVG----- 1136  
 DB 2054 -----DULRMSYCYAKF-EGGKCSSPKSRHNSKOECCALGEGMGWDP 2096  
 QY 1137 -----IYDYS-----RANKEMMDVPCYRKMDVN-----CTCKA 1166  
 DB 2097 ELCPTPEPAFRQIDCPYSGIITVGPDSAVDMDBCKEPPV--CKHGQCIINTGSRCECP 2155  
 QY 1167 GYVGDGFSQ 1175  
 DB 2156 GYIIQGNQ 2164

RESULT 10  
 NOTCH\_BRARE  
 ID NOTCH\_BRARE STANDARD: PRT: 2437 AA.  
 AC P46530:  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein precursor.  
 GN NOTCH OR NOTCH1A.  
 OS Brachydanio rerio (zebrafish) (zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=94128602; PubMed=8297791;  
 RA Bierkamp C., Campos-Ortega J.A.;  
 RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and  
 RT its pattern of transcription during early embryogenesis.";  
 RL Mech. Dev. 43:87-100(1993).  
 CC -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING  
 CC EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE  
 CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION  
 CC STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED.  
 CC ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND  
 CC NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE  
 CC ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE  
 CC AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING

CC BRAIN AND HEAD REGIONS.  
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X69088; CAA48831.1; -  
DR HSSP: P00740; 1EDM.  
DR ZFIN: ZDB-GENE-990415-173; notch1a.  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000742; EGF 2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF-IT.  
DR InterPro: IPR000800; Notch.  
DR Pfam: PF00023; ank; 6.  
DR Pfam: PF00066; notch; 3.  
DR Pfam: PF00066; notch; 3.  
DR PRINTS: PR001010; EGFBL00D.  
DR PRINTS: PR01452; NOTCH.  
DR SMART: SM00248; ANK; 5.  
DR SMART: SM00179; EGF\_CA; 19.  
DR SMART: SM00001; EGF\_Like; 16.  
DR SMART: SM00004; NL; 3.  
DR PROSITE: PS50088; ANK\_REPEAT; 4.  
DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 23.  
DR PROSITE: PS00022; EGF\_L1; 34.  
DR PROSITE: PS01187; EGF\_CA; 22.  
DR PROSITE: PS01186; EGF\_2; 28.  
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
KW Transmembrane; Signal; Glycoprotein.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHATN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.  
FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).  
FT TRANSEM 1725 1747 POTENTIAL.  
FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 21 57 EGF-LIKE 1.  
FT DOMAIN 58 98 EGF-LIKE 2.  
FT DOMAIN 101 138 EGF-LIKE 3.  
FT DOMAIN 139 175 EGF-LIKE 4.  
FT DOMAIN 177 215 EGF-LIKE 5.  
FT DOMAIN 217 254 EGF-LIKE 6.  
FT DOMAIN 256 292 EGF-LIKE 7.  
FT DOMAIN 294 332 EGF-LIKE 8.  
FT DOMAIN 334 370 EGF-LIKE 9.  
FT DOMAIN 371 409 EGF-LIKE 10.  
FT DOMAIN 411 449 EGF-LIKE 11.  
FT DOMAIN 451 487 EGF-LIKE 12.  
FT DOMAIN 489 524 EGF-LIKE 13.  
FT DOMAIN 526 562 EGF-LIKE 14.  
FT DOMAIN 564 599 EGF-LIKE 15.  
FT DOMAIN 601 637 EGF-LIKE 16.  
FT DOMAIN 639 674 EGF-LIKE 17.  
FT DOMAIN 676 712 EGF-LIKE 18.  
FT DOMAIN 714 749 EGF-LIKE 19.  
FT DOMAIN 751 787 EGF-LIKE 20.  
FT DOMAIN 789 825 EGF-LIKE 21.  
FT DOMAIN 827 865 EGF-LIKE 22.  
FT DOMAIN 867 903 EGF-LIKE 23.  
FT DOMAIN 905 941 EGF-LIKE 24.  
FT DOMAIN 943 979 EGF-LIKE 25.  
FT DOMAIN 981 1017 EGF-LIKE 26.  
FT DOMAIN 1019 1055 EGF-LIKE 27.  
-----  
FT DOMAIN 1057 1093 EGF-LIKE 28.  
FT DOMAIN 1095 1141 EGF-LIKE 29.  
FT DOMAIN 1143 1179 EGF-LIKE 30.  
FT DOMAIN 1181 1217 EGF-LIKE 31.  
FT DOMAIN 1219 1263 EGF-LIKE 32.  
FT DOMAIN 1265 1303 EGF-LIKE 33.  
FT DOMAIN 1305 1344 EGF-LIKE 34.  
FT DOMAIN 1346 1382 EGF-LIKE 35.  
FT DOMAIN 1385 1423 EGF-LIKE 36.  
FT REPEAT 1446 1486 LIN/NOTCH 1.  
FT REPEAT 1487 1520 LIN/NOTCH 2.  
FT REPEAT 1521 1561 LIN/NOTCH 3.  
FT REPEAT 1867 1910 ANK 1.  
FT REPEAT 1915 1944 ANK 2.  
FT REPEAT 1948 1978 ANK 3.  
FT REPEAT 1982 2011 ANK 4.  
FT REPEAT 2015 2044 ANK 5.  
FT REPEAT 2048 2077 ANK 6.  
FT DOMAIN 2265 2276 POLY-GLN (OPA-REPEAT).  
FT DISULFD 25 35 BY SIMILARITY.  
FT DISULFD 29 45 BY SIMILARITY.  
FT DISULFD 47 56 BY SIMILARITY.  
FT DISULFD 62 73 BY SIMILARITY.  
FT DISULFD 67 86 BY SIMILARITY.  
FT DISULFD 88 97 BY SIMILARITY.  
FT DISULFD 105 116 BY SIMILARITY.  
FT DISULFD 110 126 BY SIMILARITY.  
FT DISULFD 128 137 BY SIMILARITY.  
FT DISULFD 143 154 BY SIMILARITY.  
FT DISULFD 148 163 BY SIMILARITY.  
FT DISULFD 165 174 BY SIMILARITY.  
FT DISULFD 181 194 BY SIMILARITY.  
FT DISULFD 188 203 BY SIMILARITY.  
FT DISULFD 205 214 BY SIMILARITY.  
FT DISULFD 221 232 BY SIMILARITY.  
FT DISULFD 226 242 BY SIMILARITY.  
FT DISULFD 244 253 BY SIMILARITY.  
FT DISULFD 260 271 BY SIMILARITY.  
FT DISULFD 265 280 BY SIMILARITY.  
FT DISULFD 282 291 BY SIMILARITY.  
FT DISULFD 298 311 BY SIMILARITY.  
FT DISULFD 305 320 BY SIMILARITY.  
FT DISULFD 322 331 BY SIMILARITY.  
FT DISULFD 338 349 BY SIMILARITY.  
FT DISULFD 343 358 BY SIMILARITY.  
FT DISULFD 360 369 BY SIMILARITY.  
FT DISULFD 375 386 BY SIMILARITY.  
FT DISULFD 380 397 BY SIMILARITY.  
FT DISULFD 399 408 BY SIMILARITY.  
FT DISULFD 415 428 BY SIMILARITY.  
FT DISULFD 422 437 BY SIMILARITY.  
FT DISULFD 437 448 BY SIMILARITY.  
FT DISULFD 455 466 BY SIMILARITY.  
FT DISULFD 460 475 BY SIMILARITY.  
FT DISULFD 477 486 BY SIMILARITY.  
FT DISULFD 493 503 BY SIMILARITY.  
FT DISULFD 512 523 BY SIMILARITY.  
FT DISULFD 514 523 BY SIMILARITY.  
FT DISULFD 530 541 BY SIMILARITY.  
FT DISULFD 535 550 BY SIMILARITY.  
FT DISULFD 552 561 BY SIMILARITY.  
FT DISULFD 568 578 BY SIMILARITY.  
FT DISULFD 573 587 BY SIMILARITY.  
FT DISULFD 589 598 BY SIMILARITY.  
FT DISULFD 605 616 BY SIMILARITY.  
FT DISULFD 610 625 BY SIMILARITY.  
FT DISULFD 627 636 BY SIMILARITY.  
FT DISULFD 643 653 BY SIMILARITY.  
FT DISULFD 648 662 BY SIMILARITY.  
FT DISULFD 664 673 BY SIMILARITY.  
FT DISULFD 680 691 BY SIMILARITY.  
FT DISULFD 685 700 BY SIMILARITY.  
FT DISULFD 702 711 BY SIMILARITY.

FT DISULFID 718 728 BY SIMILARITY.  
FT DISULFID 723 737 BY SIMILARITY.  
FT DISULFID 739 748 BY SIMILARITY.  
FT DISULFID 755 766 BY SIMILARITY.  
FT DISULFID 760 775 BY SIMILARITY.  
FT DISULFID 777 786 BY SIMILARITY.  
FT DISULFID 793 804 BY SIMILARITY.  
FT DISULFID 798 813 BY SIMILARITY.  
FT DISULFID 815 824 BY SIMILARITY.  
FT DISULFID 831 842 BY SIMILARITY.  
FT DISULFID 836 853 BY SIMILARITY.  
FT DISULFID 855 864 BY SIMILARITY.  
FT DISULFID 871 882 BY SIMILARITY.  
FT DISULFID 876 891 BY SIMILARITY.  
FT DISULFID 893 902 BY SIMILARITY.  
FT DISULFID 909 920 BY SIMILARITY.  
FT DISULFID 914 929 BY SIMILARITY.  
FT DISULFID 931 940 BY SIMILARITY.  
FT DISULFID 947 958 BY SIMILARITY.  
FT DISULFID 952 967 BY SIMILARITY.  
FT DISULFID 969 978 BY SIMILARITY.  
FT DISULFID 1023 1034 BY SIMILARITY.  
FT DISULFID 1028 1043 BY SIMILARITY.  
FT DISULFID 1045 1054 BY SIMILARITY.

Query Match 6.0%; Score 473.5; DB 1; Length 2437;  
Best local Similarity 20.3%; Pred. No. 3,3e-25;  
Matches 303; Conservative 123; Mismatches 447; Indels 617; Gaps 84;

OY 135 NRCNDNDIIVRGECC-----KCSQQA-PCPLETRPLRETRKCIYSIY 176  
DB 146 NPCANG-----GQSAFESHYICTCPNFGOTCRQDVNECAVSPSCRNGCICINEV- 198  
OY 177 FMCKRSVFIGCORP-----CVRTITRACMLASLAHNAKAPRGVKKALGTASVYMGVN 231  
DB 199 -----GSTLCKCPRELYTGRPHCOR-----LYRPLCLSP-----CRSGGTCVQTS-D 237  
OY 232 GFTGTCOCGLGFNFSTACE-----TCTE-----GRYGIHCD-----QAC----- 263  
DB 238 THTTSCSLPFTGOTCTCHANNVDDCTQHACENGRCIDGINYNGCHDKHWGQYCTEDVDE 297  
OY 264 -----SCVHGRCSGPLGSGSCDDGVGRGVKCDMEITTDNC-----NGTCH----- 305  
DB 298 CELSPNACQNGTCHNTIGGFHCYCVWAGTGDCESENI--DDCSAASCAGATCHDRVAVS 355  
OY 306 -----TSANCLDP-DGKASCKCAAFRGNGTYCT----- 334  
DB 356 FPECBPHRTGLCHLDDACTSNPCOKGNSCDTNPVSGKALCTCPGAY--TGSACNODID 413  
OY 335 -----AINACTSNAGCSTKA--DCK-----RTTP-----GNR 360  
DB 414 ECLSLANPCEHGRCLNTKSPCKCLQGYEGPRGCEMDVNECKSNPCONATCIDIQIGF 473  
OY 361 VCVCAGAYTGDGIYVLEIN-----PCL-----ENHGG----- 387  
DB 474 HCICMPGY--EGVFC-QINSDDCASQPCNLNGKCIKINSFHCPCPKFSSGLCOVDDEC 530  
OY 388 -----CDRNACTQTPRQNAVCNCLPKYTGDKGKCSL-INVCILNNGGCSPPAPCNYTEQD 442  
DB 531 ASPPCNKAGACTD-GPKKYTCCECTPGFSGIH--CELDINECAS-----SP--CHYGVCR 579  
OY 443 ORI-----CTKPRYTGDAIGYRCRSIYELPKNPSTSQYFFOLOHAHVELAGRPRTVFA 498  
DB 580 DGVAFLPCDCDRPRTYGR--LCETINI-NECLSQPCRNNGTCODRNNATICTCPKG----- 630  
OY 499 PLSSSFNHEPRIKMDQOGLMSOVLRYHVVCQQLLDNLKVTTSATTLQGEPVYSISVQ 558  
DB 631 --TTGVNCEINID-------CKRKCDDYGCIDK----- 656  
OY 559 DYVFINNEAVLSSDIITSGVIVHIDKLLSPKMLLTTPMDALGRVYLQNLTTVAANHGYT 618  
DB 657 ----INGVECVCEPGYSGSMCNINIDCALPCHNGGTGIDGV-----NSFTCLCPDGF- 706

OY 619 KESKLIQDSGLSVYIDSIHPTVFWPTDKALEALPEEQDPLFNODNKDKLSYL--- 675  
DB 707 -----BDATCLSGHNHCSSNP-----CIIHSCLDQINSYRCVC 739  
OY 676 -----KFHYIRDSKALASDLPRASAKTLOGSL- SYRCTGSDIGFLPINE---OMCR 725  
DB 740 EAGMMGRNCDININECLSNPCVNGTGCKDMTSGYLCTCRAGFSGPNQMINECASNPC- 798  
OY 726 FIHGLFLPDVGAIVAGICLLLMNPFLGRCDFETTFDI PGEC---GSC-----IFTPKC 775  
DB 799 -LNGSCIDVAGKCMQML--PYTGEVCEVNLAPCSBPCKNGVCRESDFQSFSCNC 855  
OY 776 PLKSPKGVK---KKCIYNLPFRRNVEGCONL-----CTVVIOTPPROCHGF----- 820  
DB 856 PAGMGQTCEDVDINECVARNPC---TNGGYCENLNGGFQC-----RCNNGFTGALCEND 905  
OY 821 MPDQOACP--GG-----PPTPCNNRMCMCDL---Y 845  
DB 906 IDDEPBPNGSNGGYCQDVRNGFVCVCLAGFRGERCAIDIDECVASPCRNNGNCTDCVNSY 965  
OY 846 TPMQOCICIHGFNGTACEL-----CWHGRGSPD----- 873  
DB 966 T---CSCPAGFSINGEINPTPCTESSCFNGTCVVDGISSFCVCLPFTGTANCQHDVN 1021  
OY 874 -CQPRSCSEHQCODE-----GITG-----SGEC 895  
DB 1022 ECDSDRPQONGSCQDDGVTYKCTCPHGTGLNQSLRWKODSSPCKNKGSCQMOGASPTC 1081  
OY 896 LCETIGMTAASCDPTTAVFAVCT--PACSVIATCTENNNTCV-----CNLANEG----- 940  
DB 1082 QCASGWTGIVCDVPSVSGEVAARQGVSAVLCRHAGQCVDAGNTHLCRCQAGYTGYSYCQ 1141  
OY 941 -----DGITGVNDP-----CKQNNNGCAKVAKCSQKGTQV- 971  
DB 1142 EQVDECOBPNPCQNGATCT--DYLGYSCECVPRYGHANMCKSEINBCLIS-QPCQNGGCTID 1138  
OY 972 -----SCSCKKGYKGDGYSC-IEIDPCADGV-----NGG-CHEHATCRMTGPBKH 1014  
DB 1199 LVNTPYKCSQPRGTPQ--GVHCEIDIDDCSPSYDPLTGERPRCNGRCVDRV-----Gcy 1249  
OY 1015 KCECKSHYVGDVCEPPEOLPLDRCLQDNGOCNPDAS--CADLYFQDTTVGVPHLRSPFG 1072  
DB 1250 GCVCPAGFVGE--RCEBG--VNECLSD--PCDPSGSYNVOL-----INDFCECRGTG 1296  
OY 1073 ----QYKLTDPKAKEACAKAENATITATYNQLSYAKAKAYHLCASAMLSGSRAYPTTYASQ 1128  
DB 1297 YTGKRCETVEFGCKDPTPCKNGTCAVASNTKHGYICK--CQPYSGSS-----CEYDSQ 1348  
OY 1129 KCGANVVGIVDYSRANKSEMMDVFCYRMKDYNCTCKAGYVGDGFSC-----SGNLLQVLM 1184  
DB 1349 SCGS-----LRCRN-GATCVSGHLSPPCLCLAPGFSGHCQOTRM 1385  
OY 1185 SFPSLTN-----PLTEVLAFSKSSARQ 1207  
DB 1386 DSPCLVBPVNGTQOPISDAFERYCSCPANFNGLCHILIDYSPSGQGR 1435

RESULT 11  
NTC3\_MOUSE  
ID NTC3\_MOUSE STANDARD; PRT: 2318 AA.  
AC 061982;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neurogenic locus notch 3 protein.  
GN NOTCH3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ICR X SWISS WEBSTER;

RX MEDLINE=95001556; PubMed=7918097;  
RA Lardelli M., Dalstrand J., Lendahl U.;  
RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal  
RT growth factor-repeats and is expressed in proliferating  
RT neuroepithelium."  
RL Mech. Dev. 46:123-136(1994).  
CC -1- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING  
CC VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE  
CC DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.  
CC -1- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.  
CC -1- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.  
CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X74760; CA52776.1; -  
DR HSSP; P00740; 11XA.  
DR MGD; MGI:99460; Notch3.  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR000152; ASX\_hydroxyl.  
DR InterPro: IPR000561; EGF-1like.  
DR InterPro: IPR000742; EGF-2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF\_II.  
DR InterPro: IPR000800; Notch.  
DR Pfam; PF000023; ank; 6.  
DR Pfam; PF00008; EGF; 34.  
DR Pfam; PF00066; notch; 3.  
DR PRINTS; PRO0010; EGFBL00D.  
DR PRINTS; PRO1452; NOTCH.  
DR SMART; SM00248; ANK; 5.  
DR SMART; SM00179; EGF\_Ca; 19.  
DR SMART; SM00001; EGF\_Like; 13.  
DR SMART; SM00004; NL; 3.  
DR PROSITE; PS50088; ANK\_REPEAT; 4.  
DR PROSITE; PS50297; ANK\_REP\_REGION; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 18.  
DR PROSITE; PS00022; EGF\_1; 33.  
DR PROSITE; PS01186; EGF\_2; 27.  
DR PROSITE; PS01187; EGF\_Ca; 17.  
KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;  
KW ANK repeat; Glycoprotein.  
FT DOMAIN 1 1643 EXTRACELLULAR.  
FT TRANSMEM 1644 1664 POTENTIAL.  
FT DOMAIN 1665 2318 CYTOPLASMIC.  
FT DOMAIN 2242 2261 PEST.  
FT DOMAIN 39 78 EGF-Like 1.  
FT DOMAIN 79 119 EGF-Like 2.  
FT DOMAIN 120 157 EGF-Like 3.  
FT DOMAIN 159 196 EGF-Like 4.  
FT DOMAIN 198 235 EGF-Like 5.  
FT DOMAIN 237 273 EGF-Like 6.  
FT DOMAIN 275 313 EGF-Like 7.  
FT DOMAIN 315 351 EGF-Like 8.  
FT DOMAIN 352 390 EGF-Like 9.  
FT DOMAIN 392 430 EGF-Like 10.  
FT DOMAIN 432 468 EGF-Like 11.  
FT DOMAIN 470 506 EGF-Like 12.  
FT DOMAIN 508 544 EGF-Like 13.  
FT DOMAIN 546 581 EGF-Like 14.  
FT DOMAIN 583 619 EGF-Like 15.  
FT DOMAIN 621 656 EGF-Like 16.  
FT DOMAIN 658 694 EGF-Like 17.  
FT DOMAIN 696 731 EGF-Like 18.  
FT DOMAIN 735 771 EGF-Like 19.

FT DOMAIN 772 809 EGF-Like 20.  
FT DOMAIN 811 848 EGF-Like 21, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 850 886 EGF-Like 22, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 888 923 EGF-Like 23, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 925 961 EGF-Like 24.  
FT DOMAIN 963 999 EGF-Like 25.  
FT DOMAIN 1001 1035 EGF-Like 26.  
FT DOMAIN 1037 1083 EGF-Like 27.  
FT DOMAIN 1085 1121 EGF-Like 28.  
FT DOMAIN 1123 1159 EGF-Like 29, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1161 1204 EGF-Like 30, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 1206 1245 EGF-Like 31.  
FT DOMAIN 1247 1288 EGF-Like 32.  
FT DOMAIN 1290 1326 EGF-Like 33.  
FT DOMAIN 1336 1374 EGF-Like 34.  
FT REPEAT 1388 1428 LIN/NOTCH 1.  
FT REPEAT 1429 1467 LIN/NOTCH 2.  
FT REPEAT 1468 1503 LIN/NOTCH 3.  
FT REPEAT 1539 1868 ANK 1.  
FT REPEAT 1872 1902 ANK 2.  
FT REPEAT 1906 1935 ANK 3.  
FT REPEAT 1939 1968 ANK 4.  
FT REPEAT 1972 2001 ANK 5.  
FT DISULFID 43 55 BY SIMILARITY.  
FT DISULFID 49 66 BY SIMILARITY.  
FT DISULFID 68 77 BY SIMILARITY.  
FT DISULFID 83 94 BY SIMILARITY.  
FT DISULFID 88 107 BY SIMILARITY.  
FT DISULFID 109 118 BY SIMILARITY.  
FT DISULFID 124 135 BY SIMILARITY.  
FT DISULFID 129 145 BY SIMILARITY.  
FT DISULFID 147 156 BY SIMILARITY.  
FT DISULFID 163 175 BY SIMILARITY.  
FT DISULFID 169 184 BY SIMILARITY.  
FT DISULFID 186 195 BY SIMILARITY.  
FT DISULFID 202 213 BY SIMILARITY.  
FT DISULFID 207 223 BY SIMILARITY.  
FT DISULFID 225 234 BY SIMILARITY.  
FT DISULFID 241 252 BY SIMILARITY.  
FT DISULFID 246 261 BY SIMILARITY.  
FT DISULFID 263 272 BY SIMILARITY.  
FT DISULFID 279 292 BY SIMILARITY.  
FT DISULFID 286 301 BY SIMILARITY.  
FT DISULFID 303 312 BY SIMILARITY.  
FT DISULFID 319 330 BY SIMILARITY.  
FT DISULFID 324 339 BY SIMILARITY.  
FT DISULFID 341 350 BY SIMILARITY.  
FT DISULFID 356 367 BY SIMILARITY.  
FT DISULFID 361 378 BY SIMILARITY.  
FT DISULFID 380 389 BY SIMILARITY.  
FT DISULFID 396 409 BY SIMILARITY.  
FT DISULFID 403 418 BY SIMILARITY.  
FT DISULFID 420 429 BY SIMILARITY.  
FT DISULFID 436 447 BY SIMILARITY.  
FT DISULFID 441 456 BY SIMILARITY.  
FT DISULFID 458 467 BY SIMILARITY.  
FT DISULFID 474 485 BY SIMILARITY.  
FT DISULFID 479 494 BY SIMILARITY.  
FT DISULFID 496 505 BY SIMILARITY.  
FT DISULFID 512 523 BY SIMILARITY.  
FT DISULFID 517 532 BY SIMILARITY.  
FT DISULFID 534 543 BY SIMILARITY.  
FT DISULFID 550 560 BY SIMILARITY.  
FT DISULFID 555 569 BY SIMILARITY.  
FT DISULFID 571 580 BY SIMILARITY.  
FT DISULFID 587 598 BY SIMILARITY.  
FT DISULFID 592 607 BY SIMILARITY.  
FT DISULFID 609 618 BY SIMILARITY.  
FT DISULFID 625 635 BY SIMILARITY.  
FT DISULFID 630 644 BY SIMILARITY.  
FT DISULFID 646 655 BY SIMILARITY.  
FT DISULFID 662 673 BY SIMILARITY.  
FT DISULFID 667 682 BY SIMILARITY.







```
FT DISULFD 629 638 BY SIMILARITY.
FT DISULFD 645 655 BY SIMILARITY.
FT DISULFD 650 664 BY SIMILARITY.
FT DISULFD 666 675 BY SIMILARITY.
FT DISULFD 682 693 BY SIMILARITY.
FT DISULFD 687 702 BY SIMILARITY.
FT DISULFD 704 713 BY SIMILARITY.
FT DISULFD 720 730 BY SIMILARITY.
FT DISULFD 725 739 BY SIMILARITY.
FT DISULFD 741 750 BY SIMILARITY.
FT DISULFD 757 768 BY SIMILARITY.
FT DISULFD 762 777 BY SIMILARITY.
FT DISULFD 779 788 BY SIMILARITY.
FT DISULFD 795 806 BY SIMILARITY.
FT DISULFD 800 815 BY SIMILARITY.
FT DISULFD 817 826 BY SIMILARITY.
FT DISULFD 833 844 BY SIMILARITY.
FT DISULFD 838 855 BY SIMILARITY.
FT DISULFD 857 866 BY SIMILARITY.
FT DISULFD 873 884 BY SIMILARITY.
FT DISULFD 878 893 BY SIMILARITY.
FT DISULFD 895 904 BY SIMILARITY.
FT DISULFD 911 922 BY SIMILARITY.
FT DISULFD 916 931 BY SIMILARITY.
FT DISULFD 933 942 BY SIMILARITY.
FT DISULFD 987 998 BY SIMILARITY.
FT DISULFD 992 1007 BY SIMILARITY.
FT DISULFD 1009 1018 BY SIMILARITY.
FT DISULFD 1025 1036 BY SIMILARITY.
FT DISULFD 1030 1045 BY SIMILARITY.
FT DISULFD 1047 1056 BY SIMILARITY.
FT DISULFD 1063 1074 BY SIMILARITY.
FT DISULFD 1068 1083 BY SIMILARITY.
FT DISULFD 1085 1094 BY SIMILARITY.
FT DISULFD 1101 1122 BY SIMILARITY.
FT DISULFD 1116 1131 BY SIMILARITY.
```

Query Match Best Local Similarity 5.7%; Score 446; DB 1; Length 2531;  
Matches 275; Conservative 112; Mismatches 474; Indels 420; Gaps 78;

```
QY 135 NRCNDNTIIVREGCGKSOQAPCLETRKRCIYIYFMGRS--VFIGCPQ-- 190
DB 180 NCSQMP-----GLCHRGGTCHNEIGSYR---CACRATHGTGHCELPYPCSPSPC 227
QY 191 -----CYRTITTRACMLASLAHNAKRPARGEVKMKALGTASWIDVNGT 233
DB 228 QNGGTCRPTGDTTHERCACLPRFAGQNC-----EENVDDCPG--NCKNGGACV-DGVN-T 278
QY 234 GTCOCGLGFNG-----TACE---TCTEGKGIHC-----DQACS 264
DB 279 YNCRCPREPTGOYCTEDVEDVECOLMPACONAGTCHNSHGYNVCYNGMTGEDCSQNIID 338
QY 265 CVHGRCQSGP-----LGDSQCDQDVGWRGVKCDMEITTDNC--NGTCHTSANCLLD--DGK 317
DB 339 CASACFOGATGCHDRVASYFCBPGRGTLLCHL---NMACISNPNESNODTINVNGK 395
QY 318 ASCKACAGRGNGVCTA--INACETISNGGCKTADCKRTTPGNGRVGVCACTGDDGIVC- 375
DB 396 AICTCPRGY--TGPAQSOVDDEALGANPCENHAKCLNTL-GSFECQCLQGYGPR--CE 450
QY 376 LEINPLENHGGGDRNAEC--TOTGPNQAVCNCLPKYTGQGVKVSU-----INVCCLTNG 428
DB 451 IDVNECISN--FCQNDATCLDQIGEFQ--CLCMFGY--EGVYCEINTDECASPCL-HNG 503
QY 429 GSPFAFCVNTQDQRICTCKPDYTGDIYVRCGSIYGLPKPNSTSOYFFQLOEHAHREL 488
DB 504 RC-----VDKINEFLCCCPKFSGH--LCQYDV--DECASTRPC-----KNAKCL 544
QY 489 AGRGPTTVAPLSSSPNH-EPRIKMD-----QQGLM----- 519
DB 545 DGPNTYTCVCTEGYTGTHCEVDIDECDDPCHIGLCKDGVAFPTCLQPGYTGHNCEINI 604
```

```
QY 520 ----SQVLRYHYVCCOQLLDN-----LKVT-----SATTLQGEVYSIVSDTVET 563
DB 605 NECHSQPCR-HGGTQCD--RDNYLCLCLKGTTPRCNEIFNLDDCANPDCDSGTCLDKI-- 659
QY 564 NNEAKVLSSDIISTNGVIVH-IDKLL-SPKNLLITPKDALGRVLOMLTYVAANHGYTKES 621
DB 660 --DGYEACBEPYGTGSMCVNVIDECAGSPCHNGGTCEDEGIACF-----TCRCEGY---- 708
QY 622 KLIQDSGLSVITDSIHTYVTWPTDKALEALPRQDQFLFNQDKDKLSYLKRVHVR 681
DB 709 --HDPYTCISEVNECNSNP-----CHGACRGDINGT----- 737
QY 682 DSKALASDI.PRSASWKTLQSGELSVRCGTGSDIGELFLNEQWCFTHRGLEFDVGAAYGI 741
DB 738 -----KCDCAIGWS-----GTNCDLNNNECESNPC--VNGGICKD--MTSGY 775
QY 742 DCLLMNPTLQGRCDPTFTTEDIPEGC---GSCI-----FPPKCPD----- 777
DB 776 VCTCRGEGFSQPNQOTINI.ECASNPCLNOGTCIDVAGYKNCPLPYTGATCEVVLAPCAT 835
QY 778 -KSKPRGVKK-----KCIYNPLPRRNVBGCN----- 804
DB 836 SPCKNSGVCKESEDYESFSCVCPFGWOGOTCEIDINECVKSPC---RHGASCONFGSYR 892
QY 805 -LCTTVIQTTPRC-----CH-----GYFMPDCOACPG-----GPD 833
DB 893 CLCQAGYTGRCNCSDDIDDCRPNPCHNGSGCTDQVNAAFCDCLPRFGACATCEEDINECATN 952
QY 834 PCNNRGMCRDL---YTPMGOCILCHTGFNGTACELCMHGRFGPDCQPRSGSEHGQDEGIT 890
DB 953 PCQNGANCTDCVDSYF---CTCPTGFNGIHC-----NNTPDCTESSCFNGGTGCVDGI- 1002
QY 891 GSGECICETGWTASCDPTPAVFAVCTPACSVIATGTEN---NTGCNLYNEBDDGITC-T 946
DB 1003 NSFTCLCPREFGTGYCYQYD--VNECDSPCLHGCTQDSTGYTKCPCPGY--TGLNCON 1058
QY 947 VVDFCKQNNNGCAKAVKCSQKGTQVSCSKKGYGSDYCEIDPCADGVNGSCH----- 1001
DB 1059 LVHWC--DSAPCKNGKCMQNTNQYHCECRSG--TGPNC-----DVLVSCEVAQK 1107
QY 1002 -----EHATCRMTGPKKHKCECKSHYGVGVDCERBQLPDLKCLQDNQCHPDASC 1052
DB 1108 RGIQVTLLOHGGLCYDEEDKHCHCOAGYTG--YCEDE---VDEB--SPNQCQMGATC 1160
QY 1053 ADLYFQDITVGVPHLSPGLQYKILTFDKAKEACAKAPATITATYNOLSYAQKATYHLCASG 1112
DB 1161 TDY-----LGFSCKCVAGYHGSNCSEETNECLSQPC---ONGGTCLDLTFTYKQSC- 1209
QY 1113 WLESGRVAYPTTYASQKCGANVGI---VDYGRANKSEKMDVEFY-----RMKDVNC 1162
DB 1210 -----PRGTQGVHCELINVDCHPPLDPASRSK-----CFNNGTCVQDVQVGYTC 1253
QY 1163 TCKAGYVGDGFFSCSGMLQVL 1183
DB 1254 TCPRGFVGE--RECDVNECL 1272
```

RESULT 13  
NTCL\_MOUSE  
ID NTCL\_MOUSE STANDARD: PRT: 2531 AA.  
AC 001705;  
DT 01-NOV-1995 (rel. 32, created)  
DT 01-FEB-1996 (rel. 33, last sequence update)  
DT 16-OCT-2001 (rel. 40, last annotation update)  
DE Neutrogenic locus notch homolog protein 1 precursor (NOTCH protein).  
GN NOTCH1 OR MOTCH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxId=10090;  
RN 111  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;



FT	DISULFID	532	543	BY SIMILARITY.
FT	DISULFID	537	552	BY SIMILARITY.
FT	DISULFID	554	563	BY SIMILARITY.
FT	DISULFID	570	580	BY SIMILARITY.
FT	DISULFID	575	589	BY SIMILARITY.
FT	DISULFID	591	600	BY SIMILARITY.
FT	DISULFID	607	618	BY SIMILARITY.
FT	DISULFID	612	627	BY SIMILARITY.
FT	DISULFID	629	638	BY SIMILARITY.
FT	DISULFID	645	655	BY SIMILARITY.
FT	DISULFID	650	664	BY SIMILARITY.
FT	DISULFID	666	675	BY SIMILARITY.
FT	DISULFID	682	693	BY SIMILARITY.
FT	DISULFID	687	702	BY SIMILARITY.
FT	DISULFID	704	713	BY SIMILARITY.
FT	DISULFID	720	730	BY SIMILARITY.
FT	DISULFID	725	739	BY SIMILARITY.
FT	DISULFID	741	750	BY SIMILARITY.
FT	DISULFID	757	768	BY SIMILARITY.
FT	DISULFID	762	777	BY SIMILARITY.
FT	DISULFID	779	788	BY SIMILARITY.
FT	DISULFID	795	806	BY SIMILARITY.
FT	DISULFID	800	815	BY SIMILARITY.
FT	DISULFID	817	826	BY SIMILARITY.
FT	DISULFID	833	844	BY SIMILARITY.
FT	DISULFID	838	855	BY SIMILARITY.
FT	DISULFID	857	866	BY SIMILARITY.
FT	DISULFID	873	884	BY SIMILARITY.
FT	DISULFID	878	893	BY SIMILARITY.
FT	DISULFID	895	904	BY SIMILARITY.
FT	DISULFID	911	922	BY SIMILARITY.
FT	DISULFID	916	931	BY SIMILARITY.
FT	DISULFID	933	942	BY SIMILARITY.
FT	DISULFID	987	998	BY SIMILARITY.
FT	DISULFID	992	1007	BY SIMILARITY.
FT	DISULFID	1009	1018	BY SIMILARITY.
FT	DISULFID	1025	1036	BY SIMILARITY.
FT	DISULFID	1030	1045	BY SIMILARITY.
FT	DISULFID	1047	1056	BY SIMILARITY.
FT	DISULFID	1063	1074	BY SIMILARITY.

Query Match 5.7%; Score 445; DB 1; Length 2531;  
Best Local Similarity 22.6%; Pred. No. 3.7e-23;  
Matches 288; Conservative 104; Mismatches 477; Indels 404; Gaps 84;

QY	135	NCNDNDITIVGECGCKSQQ-----APCPLETKPLRTKRCYISYFMGRKSVFIGQ	188
DB	180	NECSNPFGLCRHG--GHCHNEIGSYRCACCAHTGPHCEL-----PYVPCS	223
QY	189	PC-CVRTIITR-----ACMLASLA-----HNKKPAPGEVKKMCAIGTASVMDGVNCT	233
DB	224	PSPCONGACRPTGTGTTTHCAC-LGCFAGCNCEENVDCPG--NNCKNGACV-DGVN-T	278
QY	234	GTCGGGLGFGN-----TACE--TC--TEGKYIHC-----DQACS-----	264
DB	279	YKRCRPVEVTGYCTEDVDECOLMPNACONAGCHNTHGYNOCVYNGWTGECSENID	338
QY	265	CYHGKCSQGP-----LGDSGCCDVGWRGVKCDMETTTNCGTGTTSANCLIDP-DGKA	318
DB	339	CASACFCQDGTCHDRVASFCECPHGRTGLCHLKHAC--ISNPCEGSCNDTNPNYNGKR	396
QY	319	SCKCAAGFGNGTVCJA-INACETSNNGCGSTKADCKRTTGNRNVCCYCKAGYTGDIIVC-L	376
DB	397	ICTCSBG--TGPACSDYVDECDLGNRCENHAGKCLNTL-GSEFCCLDGYTGP-G-CEI	451
QY	377	EINPLENHGGCDRNAEC-TGTGPNQAVNCLPKYTGDKVCSL-----INVCLTNNGG	429
DB	452	DVNECISN--PCONDATCLDQIGEPD--CICMPGY--EGVYCEINTDECASSPCL-HNH	504
QY	430	C-----SPRAF-----CNYTED-----ORICTCKPDYTGDIIV	458
DB	505	CMDKIHFOCQCPKGFNGHLCQYDVDECASTPCKNGAKCLDGPNTYTCVCTEGYTGTH--	562

QY	459	CRGSITGELPKNESTQYFFLOEHAHARELAGRPETVEAPLSSPNH-EPRIKMDQOG	517
DB	563	CEVDI--DEDDPDPG-----HGSCKDGAFTTCLQGPYTHHETINNE-----	606
QY	518	LMSQVLRHYHVGCQQLLDN-----LKVT-----SATTLQGPVSVSVODYINN	565
DB	607	CHSQPCR-HGHTQD--RNSYLCCLKTGTGENCEININDCASNPCDSGTCLDKI----	659
QY	566	EAKVLSDDIISNGVIHV-IDKLL-SPKNLLITPKALGRVLONLITVAANHGYTFSKL	623
DB	660	DGYECACEPGYVGSNCNVNIDECAGSPCHNGGCEGDIAGF-----TCRCPEGY-----	708
QY	624	IDDSGLSVITPISHPRVVFWPTDKALBALPPEODPLFNQDNKDKLSYLFKHVIRDS	683
DB	709	-HDPTLSEVNECSNP-----CINACRCGLNGY-----	737
QY	684	KALASDLPRASAKTLOGSELVYRCGTGSDIGELFLNEQMCREFIRHGLLFVDGAVAYIDC	743
DB	738	-----KCDAPGWS-----GTNCDINNNECESNPC--VNGGTCKD--MTSGYVC	777
QY	744	LMNPTLGRCDTFTTFDIPGEC--GSCl-----FTPKCPL-----K	778
DB	778	TCREGFSGPNCQTNINECASNPCLNGTCTIDVAGYKCNCPLYTGATGEVVLAPCATSP	837
QY	779	SKPKGVK-----KCIYNPLFPRRVNCGCN-----L	805
DB	838	CKNNGVCKSEDEYESTSCVCPHMOGQCEVINDNECVKSPC--RHGACQQNTNGSYRL	894
QY	806	CTVNIQTPRC-----CH-----GYFMPDQACPG-----GPDPC	835
DB	895	COAGYGRMCESDIDDCRPNCHNGSCTDINTAFCDCLPGFGAFCEEDINECASNPC	954
QY	836	NNRGMKRDL---YTPMGQCLHTGFNGTACELCMHNRFPDQPRGCSHGGCDEBITGS	892
DB	955	QNGANCTDQVDSY---CTCPVGFNGIHCE-----NNPRTCESSCFNGGTCVDGI-NS	1004
QY	893	GECICETGWTAAACDPTFAVFAVCTPACSVHATCTEN--NTCVCMANVEGDITC--TYV	948
DB	1005	FTCLCPGFTTGSTCQYD--VNECDSRPLAGGTQODSYGTACTCPQGI--TGLMCONLY	1060
QY	949	DFCKQNNGCAYAKVACSQKTOVSCSKKRGKD-----GYSCIEIDPCADGVNGG--CH	1001
DB	1061	RMC--DSAPCKNNGRCWMQNTQYHCECRSGMTGVNCDVLVSVG--EVAAGKRGIDVLLQ	1117
QY	1002	EHATCRMTGPKHKKCEKSKHYVGDVDCPEBOLPLDRCIODNGQCHPRDASCADLTFDPTT	1061
DB	1118	HGGLCYDEG-DKHVCHCQAGYTGSS--YCEDE--VDEC--SPNPCONGATCTDY-----	1163
QY	1062	VGVFHLRSPDLOY-KLTFPKAKKACAKKAATJATYNQLSYAOAKAYHLCASGMLDSGRVA	1120
DB	1164	LGFSCKVYAGTHGNSCSEINECLSOPCONGSTCIDLTINSYK-----CSC-----	1209
QY	1121	YPTTVAOKCGANVGI---VDYGRANKSEMDVFCY-----RMKDVCNCTCKAGYVG	1170
DB	1210	PRGTGVHCEINVDCHRPDPASRSPK-----CFNNGTCVDQVGVYTTCTCPGFGV	1261
QY	1171	DGFSCGNLLQYL 1183	
DB	1262	E--RCGDVNECL 1272	

RESULT 14  
NTC4\_MOUSE  
AC P31695; 062389;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neurogenic locus notch homolog protein 4 precursor (Transforming protein INT-3).  
GN NOTCH4 OR INT3 OR INT-3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;







FT	CARBOHYD	3965	3965	N-LINKED (GICNAC. .) (POTENTIAL).
FT	CARBOHYD	4140	4140	N-LINKED (GICNAC. .) (POTENTIAL).
FT	VASOSPILIC	1	3616	MISSING (1IN ISOSFORM XB-SHORT).
FT	CONFLICT	135	135	G -> GEOG (1IN REF. 2).
FT	CONFLICT	4038	4038	P -> G (1IN REF. 3 AND 5).
FT	CONFLICT	4163	4163	M -> I (1IN REF. 3 AND 5).
FO	SEQUENCE	AA: 4289	AA: 4644	MM; 846557C12C6556470 CMC64;

Query Match	5.58;	Score 435;	DB 1;	Length 4289;
Best Local Similarity	23.18;	Pred. No. 3.8e-22;		
Matches 223;	Conservative 71;	Mismatches 318;	Indels 354;	Gaps 58;

Search completed: June 17, 2002, 12:35:32  
Job time: 308 sec

```

OY 1047 HPDMSA--DLYFPDPTVGVGFIHLRSPLGOKKLTFDKAKAKAC--AKENAAITATYNOLSYAQ 1102
Db 730 H-DSCSCVKCKDY-----AGEDCGEARVPSASATDQGLAP 764
OY 1103 KAKYHL 1108
Db 765 GOEYOV 770

```

QY	186	GCOPCCVATITTRACMYASTASHANAKPARPENV---KKCATGTA5VMDGVNGTGTCCGCGG	242
Db	116	GLKRCQ-----TGGCCGAS---AQAQTGTDTVRLTICSL--HGEVD--LSRGTCSCEGWM	162
QY	243	NGTACETCTE-----GKYLHCDOAC5CVHGRCSOGRLDGSDDCDDGVMRGVYK	290
Db	163	GGPCTSPDPTDEIRPPSSPPSASGSCPDCCNDGRCVGRVCY-----CFPGYTGPS	212
QY	291	CDMETTDNCKNGTCHTSTANCLLDPDQKASKCAAGRGNGVYCTALINCEST5NGCCSTKA	350
Db	213	CGW----PSCGGDQGGRGRCV-----QGVYVCRAGE--GGPCCSO-RSCPR---GCSORG	257
QY	351	DKCRRTTGNRNVCAKAGYTDGIVYCLEINPCLNEHHGCCORNAECTQTGPNQAVNCMLPKY	410
Db	258	RCE----QGR-CVYDDPRTYTD---CGMRSCPR---GCSQGRGC-----ENKRCYCNPGY	301
QY	411	TGDKVCSLINVCLTNNGCSPPAFQCNYTEEDQRICTCKPDYTDGIVYGRGSIYELPKN	470
Db	302	TGED--CG-VRSQ---PRCSSORGRC---KDR-CVCDPRTYGEDCCGRSC-----	342
QY	471	PSTQYFFQQLQEHAVRLACGPFTYFAPRLSSFNHPRIKMDQOGLMSQYLRYHVVGC	530
Db	343	-----PMD-----CGE	348
QY	531	QQLLDNLKVTTSATTLQGEPSVISVSODTVFINNEAKYLLSDI1STNGVHHVIDKLSP	590
Db	349	GGRCVYDGCRCVCMPGYT--GEDCSHRT-----	372
QY	591	KNLLITRKDALGRVLQWLTVAANHGYTKSKLIDSGLLSYTD-----SIHTVTVYFW	645
Db	373	-----CPDGDGRCRCDEGCICDTGYSG---DDCGVRSRPGDNDGRGCEIDRCVCM	422
QY	646	P-----TKALEALPRPQODEFLFNODNKKDKLSYLKPHVLRD5KALASLDPR5ASKTLQGG	701
Db	423	PQYTGTDGGSNAC-----RDCRGR-----	443
QY	702	SELSRCGTGSGDIGELFLNBOMCRFIHRGLFLPDVG--AYGIDCLLMP1LTLGRCDDTPTTF	760
Db	444	-----RCENGVCY-----CNAGYSG--EDCGVRSRPGDC-----RGRGRCESGRCM	482
QY	761	DIREGCSGCIITPRCPLKSNPKG--YKKKCIYNP-----LFFRRNVECO--NLCTVVI	810
Db	483	CMPTGTGIDGCGTRACPDCKRGKRGICVDCRCVCMPTGTGEDCCSRRCRPGDGRHGHGCEJGV	542
QY	811	QTPROCHYFMPDC--OACRGPGRDPCNNRGMCRDLYTTPMGQCLCHTGNTFACELCMWG	868
Db	543	CV--CDAGYSGEDCSSTFRSGG---CRRGQGLD-----GRVCYEDVYSGEDC-----	585
QY	869	RFGPDOPRSGSEHGQDCEG1TSGSELCTGCTMTAASCTP1PAVAVCTIPAC5VHATGTE	928
Db	586	--GVRQCPNDCSQHGVCQDGV--CICMEGV5SEDCSIRT-----CP5NHGGRG-C-E	632
QY	929	NNTCVCN1YEGDGICTTVDFCKONNGGCAKAYAKCSOKG--TOYSCSKKKYCKEDG5C	986
Db	633	EGRCCLCDPGYGP--TC-AIRMCPR-----ADCRGGRGVQGVGLCHVGTGE--DC	678
QY	987	TEIDPCADGVNGSGCHEHATCRMTGCRKHKCECKSHYVGDGVDCERPOLPRLDRC1LDNQGQ	1046
Db	679	GOEFPFASACGGGCGPRELDR-----AGQCVCEGF--RGRDCA1DGTGSD--CGRREC	729

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:25:54 : Search time 19.98 Seconds  
(without alignments)  
1749.401 Million cell updates/sec

Title: US-09-842-930A-2

Perfect score: 7861

Sequence: 1 SLPSLTLRLQMPDYSIFRG.....WGHGPDMSQATTVTPR 1431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_AA:\*  
2: /cgn2-6/ptodata/2/1aa/5A.COMB.pep:\*  
3: /cgn2-6/ptodata/2/1aa/5B.COMB.pep:\*  
4: /cgn2-6/ptodata/2/1aa/6A.COMB.pep:\*  
5: /cgn2-6/ptodata/2/1aa/6B.COMB.pep:\*  
6: /cgn2-6/ptodata/2/1aa/6CTOS.COMB.pep:\*  
6: /cgn2-6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	486	6.2	2523	1	US-08-185-432-18 Sequence 18, Appl
2	480.5	6.1	2556	1	US-08-083-590A-20 Sequence 20, Appl
3	480.5	6.1	2556	3	US-08-532-384-20 Sequence 20, Appl
4	475	6.0	2703	1	US-08-185-432-19 Sequence 19, Appl
5	471.5	6.0	2556	1	US-08-185-432-17 Sequence 17, Appl
6	450.5	5.7	2471	1	US-08-185-432-16 Sequence 16, Appl
7	450.5	5.7	2471	1	US-08-083-590A-19 Sequence 19, Appl
8	450.5	5.7	2471	3	US-08-532-384-19 Sequence 19, Appl
9	413.5	5.3	1404	2	US-08-400-159-2 Sequence 2, Appl
10	413.5	5.3	1404	2	US-08-611-729A-2 Sequence 2, Appl
11	410	5.2	1193	2	US-08-400-159-10 Sequence 10, Appl
12	410	5.2	1193	3	US-08-611-729A-10 Sequence 10, Appl
13	399.5	5.1	1010	4	US-08-882-046-7 Sequence 7, Appl
14	399.5	5.1	1036	4	US-09-068-740A-6 Sequence 6, Appl
15	399.5	5.1	1187	4	US-09-068-740A-7 Sequence 7, Appl
16	399.5	5.1	1218	4	US-08-400-159-6 Sequence 6, Appl
17	399.5	5.1	1218	3	US-08-611-729A-6 Sequence 6, Appl
18	399.5	5.1	1218	4	US-08-882-046-2 Sequence 2, Appl
19	399.5	5.1	1218	4	US-08-882-046-2 Sequence 2, Appl
20	399.5	5.1	1218	4	US-09-214-278-7 Sequence 7, Appl
21	392	5.0	1219	4	US-08-882-046-5 Sequence 5, Appl
22	368	4.7	3111	2	US-08-125-077-4 Sequence 4, Appl
23	368	4.7	3111	2	US-08-125-077-4 Sequence 4, Appl
24	348	4.4	4544	1	US-08-469-486-52 Sequence 52, Appl
25	348	4.4	4544	2	US-08-469-486-52 Sequence 52, Appl
26	343	4.4	1148	4	US-08-882-046-4 Sequence 4, Appl
27	342.5	4.4	1065	2	US-08-400-159-8 Sequence 8, Appl

28	341.5	4.3	1055	4	US-09-214-278-2 Sequence 2, Appl
29	341.5	4.3	1212	4	US-09-214-278-3 Sequence 3, Appl
30	341.5	4.3	1257	3	US-08-611-729A-8 Sequence 8, Appl
31	339.5	4.3	1248	4	US-08-882-046-6 Sequence 6, Appl
32	337.5	4.3	3075	2	US-08-460-309-5 Sequence 5, Appl
33	337.5	4.3	3075	2	US-08-125-077-5 Sequence 5, Appl
34	331	4.2	1238	4	US-09-214-278-5 Sequence 5, Appl
35	325	4.1	2199	5	PCT-US95-11684-2 Sequence 2, Appl
36	324.5	4.1	1810	5	PCT-US95-11684-4 Sequence 4, Appl
37	322	4.1	1525	3	US-09-191-647-2 Sequence 2, Appl
38	322	4.1	1525	4	US-09-540-245A-2 Sequence 2, Appl
39	322	4.1	1525	4	US-09-540-153-2 Sequence 2, Appl
40	279	3.5	713	3	US-08-872-855-5 Sequence 5, Appl
41	278.5	3.5	830	3	US-08-872-855-11 Sequence 11, Appl
42	276.5	3.5	833	1	US-08-264-534-6 Sequence 6, Appl
43	276.5	3.5	833	1	US-08-083-590A-2 Sequence 2, Appl
44	276.5	3.5	833	1	US-08-465-500-6 Sequence 6, Appl
45	276.5	3.5	833	2	US-08-346-126-6 Sequence 6, Appl

#### ALIGNMENTS

RESULT 1

US-08-185-432-18

Sequence 18, Application US/08185432

Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon

APPLICANT: Busseau, Isabelle

APPLICANT: Diederich, Robert J.

APPLICANT: Xu, Tian

APPLICANT: Matsuno, Kenji

TITLE OF INVENTION: DELTA PROTEINS, NUCLEIC ACIDS, AND

NUMBER OF SEQUENCES: 23

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESS: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/185,432

FILING DATE: 21-JAN-1994

CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2523 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-18

Query Match 6.2% Score 486; DB 1; Length 2523;  
Best Local Similarity 21.9%; Pred. No. 8e-31;  
Matches 302; Conservative 131; Mismatches 457; Indels 486; Gaps 93;

```

QY 105 VYNEAPINVT-----NVAIDKGVHGLEKYLEIQRNCDNDITIVRGEQG 150
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 FICHPGVFTDKVCLTPVDNACVNNPCNNGSTCELLNSVTE-YKRCRPPGWT-----GD-- 137
QY 151 KCSQAQPCPLETKPLRETRKCI-YIYFMGRSVFIGOP----- 189
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 SCQAADPC--ASNPCANGKCLPEIYICK-----CPPEFHGATCKODINECSQNPCK 189
QY 190 ---OCVRIIRACMIAS--IAHN-----AKPAPEGVKKACALGATASVMDGVNGTGTQ 237
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 NGGQGINFEYSRGTCCQNRFTGRNDEPYVPCNPS-----C-----LNG-GRCR 233
QY 238 -----CGLPFGNACE-----TCTEG--KYGIHC-----DQAC-- 263
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 QTDPTSYDCTCLPGFSQNCENIDDCPSNNCRMGTCVDEVTNYNCCQPPDMTGQYCTE 293
QY 264 -----SCVHGRCSQGPLGSDSCDDVGMRGVRCMDMETITDNC-NGTCHTSANLL 312
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 DVDECQILMPNACQNGGTCHNTYGGYNCVNGWTEGDESENI--DDCANAACHSGATC-- 349
QY 313 DPDGKAS--CKCAAGFRNGTGTCTAINACETISNGCSTKADCKRTP--GNRVCVCKAGY 368
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 -HDIRASFYCCPHG--RTGLLCHLDNMC--ISNPCNNGSNCD-TNPVNGKALICTCPPGY 403
QY 369 TG-----DGIYC-LEINPCLENHGGCDRNAECTOT-GRNQAVCNCLPYTGDGKVCSL-I 420
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 404 TGPACNNVDYDCSLGANPC--EHGG-----RCTNLTGSGFO--CNCPOGACPR--CEIDV 452
QY 421 NVCLTNGGCSRFACNCTEEDORICTCKPRYTGDGIYCRSIVELRKNP----- 471
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 453 NECLSN--PCONDSTC-LDQIGEFQICMPG--EGLYCETNI--DECAENPCLNHGKCID 506
QY 472 -----STQYFPOLOEHAHVE-----LAGPFPYFAPLSSSFNH----- 506
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 KINERCCCPGFSGNLQHOHDECTSPCKMGAKCIDGPNYTCQCTEGFGHRCEDI 566
QY 507 EPRIKMDQOGLMSOVLRYHVVCO-----QLLDNLKVTTSATLQGEPPVS-----I 554
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 567 NECIPDPHYGCKDGIATFTCLCRPGYGRLCDNDINECLSKPOLNGQOCRENGYIC 626
QY 555 SVSOTYVFINNAKY--LSSDIISNGYIHTD-----KLSPKLLI----- 595
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 627 TCPKGTGVECTKIDDCASNLCDNGKCIDKIDYECTCEPGYTKLC--NININECS 683
QY 596 -----TPKDALGRVNLTTVAANHGYTKFSKLIDSGLSVITD-----SIHT 639
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 684 NPCRNGGCKDOI-----NGFTCVCPDG-----HDMCISEVNECSNPCIHGACID 731
QY 640 PVTVF-----WPTDKALEALPPEODFLFNQDNKDKLSTYLKRVHVRDSKALASDLPR 692
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 732 GVNKGKDCDEAGW-----SGSNCD-----INNNECESPCMN 763
QY 693 SASWYTLGSGSELVRCGTG-----SDIGELFLNBOCMRFTHRGLLPGVAAYGIDCLL 745
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 764 GGTCDMTGAYICT--CKAGFSGPNCOTNINECSSNP--C--LNHGTCIDDAVAGYKCNML 818
QY 746 -----MNPIL-----GRC--DTFTTPI--PG-----ECSCITFP-- 773
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 819 PYTGAICEAVLAPCAGSPCKNGRCKESDEFTEFSCCPGMOGOTCEIDMNECVNRPOR 878
QY 774 -----KCPKLSKPKVGVKKCIYNPLPFRNRVGECONLCTVITQTPRCH-- 817
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 879 NGATCONTNGSYKC--NCKRPGYGRNC-----EMDIDDCQ-----PNPCNNGS 920
QY 818 ---GTFMPCQACPG-----PDTPCNNRGMCRDL---YTPMGOCILHTGN 858
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 921 CSDDGIMFECN--CPAGFRPKCEEDINECASNPCKNGKANGANTCVCNSYT---CTQCPGHS 975
QY 859 GTACELCWHGRGPDOPSCSEHGOCDGIGTSGECLTETGTAASCDPTAVFAVATP 918
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 976 GJHCE-----SMTPDCTESSCENGCTIDGT-NTFTCCQCPREFGTGYCQHD--INCDCK 1027
QY 919 ACSVHATCTEN--NTCYCNLNEYEGDITC-TVVDFCKONNGCAKAVAKCSOKGTQVSGS 974

```

```

Db 1028 PCLNGTQDSDSYGYTKCTCPGY--TGLNCONLVRWC--DSSPCKNGGCKMOTNNFTACE 1083
QY 975 CKKGKYG-----DGYSCEIDPPCADGVN--GGCHEHATCRMTGPKHKECKESHVYDGV 1027
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1084 CKSGMTGYCCDVPYSVC--EVAAKQOGVDIYHLCRNSGMCVDTG--NTHRCROQAGYTS-- 1139
QY 1028 DCEBPOLPLDRCLQDNGCCHPDASCADLYFODTTGYVHLSPLQGYKLTPDKAKEACAK 1087
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1140 YCEEQ---VDEC--SPNCPQNGATCTDY-----LGGY----- 1166
QY 1088 EAATLATYNOISYAOKAK---YHLCASG---WLESGRVAYPTTASOKGANYVGIYD 1139
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1167 SCECVAGHYGVNCSSEINECLSHPCQNGGTCTIDLINITYKCSQPRGTQGVHCEINWDDCTP 1226
QY 1140 YGSRANKSEMDFV-----CY-----RMKDVNCTKAGYGVDFSCGNLLOYL 1183
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1227 F-----YDSFTLEPKCFNNGKCIDRVGGYNICICPPGFVGE--RCEGDVNECL 1271

```

## RESULT 2

US-08-083-590A-20

Sequence 20, Application US/08083590A

Patent No. 5786158

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie &amp; Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/083,590A

FILING DATE: 25-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MISTOCK, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 2556 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-083-590A-20

## Query Match

Best Local Similarity

Matches 269; Conservative 120; Mismatches 445; Indels 407; Gaps 75;

6.1%; Score 480.5; DB 1; Length 2556;

21.7%; Pred No. 2.3e-30;

```

QY 127 EYVLEIQRNCDNDITIVRGEQ--GKCSQAQPCPLE-----TKPLRETR----- 169
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 ENIDDCPGNNCKNG-----GACVDGVNTYNNCPPEMTGYCTEDVDCQILMPNACQNG 309
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 -----KCIYSIYFMGRSVFIGOPCVRTI---ITRACWLASLAHNAK----- 210

```

```
Db 310 GTCHNTHGYNVCVNGMTGE-----DCSENIIDDCASACRHGATCHDRVASFYCE 360
QY 211 -PAPGEVMMCALGTASV-----WDGVNGTGTCCGGLGFGNTAC---ETCTEGR 254
Db 361 CPHGRTGLLCHINDACISNPCNEGSMCDTNPVNGKAICTCPGSGYGPACSGQVDECSLG- 419
QY 255 YGICHDQACSCVHGRCSGPLDGSDDCDVGRGWKCDMEITTDNC-NGTCHTSANCLLD 313
Db 420 -ANPCEHNAKCTIN-----TLCSFECCLQGYTGPRC--EIDVNECVSNPCNDATC-LD 469
QY 314 PDGKASCKCAAGFRNGTVCTAINA--CETS---NGCSTKADCKRTTPGNRVCCKAG 367
Db 470 QIGFEQCMCMPEYEG---VHCEVNTDECASSPLHNRCIDKIN-----EFCECPG 519
QY 368 YTGDDIVC-LEINPCLENHNGCDRAECTGTGTNPNAVNCMLPKRYTGDKVCLINCLTN 426
Db 520 FTGH--LCQYDDECAST--PCKNGAKCLD-GPNFTYTCVTEGYTG-----THCEVD 566
QY 427 NGGCSPAFCNTEQDORI---CTCKPDYTGDIYCRGSIYELPKNPSTSQYFFLOE 482
Db 567 IDECDPDP-CHTGSCKDGVATFTCLCRGYTGH--CETINI-NECSSQPCRLRGTCQDDP 622
QY 483 HA-----VRELAGPGRFTVFAPRLSSFNHPRIKDWDQGLMSQVLRHVHVCQQLLDN 537
Db 623 NAYLCFCLKGTGP-----NCEINLDD-----CASSPCDS 652
QY 538 LKVTTSATTLGGERYSISVSQDTFVINNEAKYSSDIISTNGVIVHIDKLSPKULLTP 597
Db 653 ---GCLDLKIDYECACERGTGSMCNSIDECAGNPGCHNG------TC 693
QY 598 KDALGRVLQNTLVVAANHGTKEFSKLIDSGLLSVITDSIHPTVTVFWPTDKALEALPPE 657
Db 694 EDGI-----NGFTGCPREGY-----HDPTCLSEVNEGNSMPCV- 726
QY 658 QODPLFNODNKKDLKSYLKFFHV-----IRDSKALASDLPRSASKWTLQSGSELSVR 707
Db 727 ---HCACRDSLNGY-KCDCDPCWMSGTINCIDINNECESNPGCVNGCTCKDMT-SGIYCT 778
QY 708 CGTG-----SDIGELFLINBQMRFIHRGLLFYGVAVAGIDCLLMNPTLGRCTFTTF 760
Db 779 CREGSGNGNCOMTINECASNP--C--LNKGTCIDVAGIKCNCLL--PYTGATCEVILAP 832
QY 761 DIPGEC--GSC-----IFTPKCPILKSKPKGVK-----KKCIYPLFLFRNVEGC 802
Db 833 CAPSPCRNGGEGROSEDESFSCVCP---TAGAKGTCCEVDINECVLSPC---RHGASC 885
QY 803 ONLCTVVIOTPRCHGCTMPDCQACPRGPD-----TPCNNRGMCRD-LYTPMGOC 851
Db 886 QN-----THGYRCHQOAGYSGHNCETDIDDCRPNCHNGSGCTDGINTAFC-C 932
QY 852 LCHTGFNGTACE-----LCWHGFRG-----PPCOPRS 878
Db 933 DCLPGFRGTFCCEEDINECASDPCRNAGNCTDCVDSYTCCTCPAGFSGICHENNTPDCTESS 992
QY 879 CSEHQCDGEGITGSGECLCTETGWTAAASCDTPPAVAVC-TPACSVNATCTEN---NTCVC 934
Db 993 CENGSTCYVDGI-NSFTCLCPREGFTGSYQ---HVVNEEDSRCLLGITQDORGLHRCIC 1048
QY 935 NLNTEGDDITC-TVYDFCKONNGGCAKAYAKCSQKTOVSCSKKQYK-----DGYSCIE 988
Db 1049 POGYGP--NQONLVHMC--DSSPCKNGGKCMQWOTIYRCECPSGMTGLYCDVPSVSC-E 1103
QY 989 IDPCADGVNGG--CHEHATCRMTGPKHKHCKECSHYVG---DGYD-CEPBDLPLDRCLQ 1041
Db 1104 VAAHQGVAVARLCOHGLCYADAG-NTIHRCQOAGTGSYCEDLVDECS- 1152
QY 1042 DNGQCHPADASCADLYFODTTVGVFHLRSLPGYKLTLPDKAEKAEKAAATIATYNQLSYA 1101
Db 1153 --SPQNGATGCTDY-----LGGYSCKCVAGYHGVNCSFEIDECLG- 1190
QY 1102 OKAKHILCSAGLWESGRAVPTTYASQKCGANVGI-----VDGSRANKSEM 1149
```

```
Db 1191 -----HPCQNG---GTCLDLPNTYKC-SCPRGTQGVHCEINVDNCPVPDVSRSRK--- 1238
QY 1150 KDVFCY-----RMKDVNCTCKAGYVSDGFSQSLQVL 1183
Db 1239 ---CFNNGTCVDQGVGYSCTCPGRVGE--RCEGVNECL 1273

RESULT 3
US-08-532-384-20
; Sequence 20, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-532-384-20

Query Match 6.1%; Score 480.5; DB 3; Length 2556;
Best Local Similarity 21.7%; Pred. No. 2.3e-30;
Matches 269; Conservative 120; Mismatches 445; Indels 407; Gaps 75;

QY 127 EKVLEIQNRCNNDDTIYRGEC--GKSGQAPCLE-----TYPLRETR----- 169
Db 256 ENIDDCPGNNCKNG-----GACVDQVNTYNCPCPEWTGYCTEDYDECQLMPNACQNG 309
QY 170 -----KCIYSIYFMGRKRSVFICGCPQCVRII---ITRACMLASLAHNAK----- 210
Db 310 GTCHNTHGYNVCVNGMTGE-----DCSENIIDDCASACRHGATCHDRVASFYCE 360
QY 211 -PAPGEVMMCALGTASV-----WDGVNGTGTCCGGLGFGNTAC---ETCTEGR 254
Db 361 CPHGRTGLLCHINDACISNPCNEGSMCDTNPVNGKAICTCPGSGYGPACSGQVDECSLG- 419
QY 255 YGICHDQACSCVHGRCSGPLDGSDDCDVGRGWKCDMEITTDNC-NGTCHTSANCLLD 313
Db 420 -ANPCEHNAKCTIN-----TLCSFECCLQGYTGPRC--EIDVNECVSNPCNDATC-LD 469
```

Qy	314	PDGKACKCAAGRGVGTCTAIIIN	-CETS-----NGGCSIRKADCKRTTPGNRPVCYKAG	367
Db	470	QITEEQCMCPTEG----	VHCEVNTDECASSPCLHNGSLDKIN-----EPQCEPRTG	519
Qy	368	YTGDDGIVC-LEIPLCEIHHGGCDRNAECTOTGPNQAVCNLPKXYTGDKVCSLIINCLTN		426
Db	520	FTGH--LCQYDDECAST--PCKNAKCLID	-GPNITYCTVCTEGYTG-----THEEVD	566
Qy	427	NGGCSPPAPCNTIEDQRI----	CTCKPRYTGDIIVCGSITYGELPRKPNSTSYFPQDE	462
Db	567	IDBCDDP-CHYGSCDVGATFCTCLRPYTGHN--	CETNI--NCCSSOPCLRGTCDPD	622
Qy	483	HA-----VRELAPGPFTVAFAPLSSSFNHPRIKIMDQGLMSQVLRYHVUGCQQLLDN		537
Db	623	NAIYCLCKTTP--	-KCEIMLD-----CASSPDS	652
Qy	538	LKVTYATTLQGBPVSISVSODTFIINNEAKVLLSSDIISTNGVIHVIDLKLSPRNLLTP		597
Db	653	---GTLCDKIDGECACBEGYTGSMCNISIDECAGNPCHHG--	-----TC	693
Qy	598	KDALGRVLOULTVVAANHGYTKFSKLIQDSGLLSTITSIHPTVYVFWPTKALEALPE		657
Db	694	EDGI-----NGFRCRPEG-----HPTCLSEVNECNSPCV-	-----	726
Qy	658	QOQFLFMQDKDKLSYLFHV-----	IRDSKIASLDPRSAMKTLQSELSVR	707
Db	727	-----HGACRDSLNG-KCDDCPGMSGTNCIDINNECESNCPVAGCKRMT--SGIVCT		778
Qy	708	CGTG-----SDIGFLFNEOMCFRINHGLFDVGAVAGIDCLLMPTLQRCDDTFPTF		760
Db	779	CREGFSGPCNQOTINIECASN--C--LNNGTCIDVAGIKCMLL--PYTATEEVYLA		832
Qy	761	DIEPEC---GSC-----IFPKCPKLAKPRGVK-----KKCIAPLPRFRNREGC		802
Db	833	CAPSPCKNGEBCROSEDEYSEFSCVP----	TGAKGQICEVDINECVLSPC---RHGASC	885
Qy	803	QNLCTVVIOTPRCHGHEYFMPDCACPGPD-----	TPCNNRGMCRD-LYTPMGQC	851
Db	886	QN-----THGGYRCHCOAGAGSRNCETDIDCPRNCHNGSGCTGINTAF--C		932
Qy	852	LCHTGFNGTACB--	-----LCHNGRG-----EDCPRS	878
Db	933	DCLRGFGTGCCEEDINECASDPRCANGACTDVSYTCTCPAGFSGIHCENNTDPCRESS		992
Qy	879	CSEHGQDEGTGSGECLCTGTWATASCDTPALFAVC--TPACSVHATCTEN--NTCVC		934
Db	993	CFNGCTCVDGI--NSFTCLCPRTGTSYQ--	HVNECDSPRCLLGGTCOORGRLHRTCT	1048
Qy	935	NLANTEGDITC-TVVDFCKONNNGCAKVAKCSOKTGVSCSKKGYK--	-DQYSCIE	988
Db	1049	POGTGTP--NQMLVHMC--DSSPCKNGKCMQHTIYURCECPSGMTGLYCDVPSVSC-E		1103
Qy	989	IDPCADGVNGS--CHENATORMTGRPKHNCCEKSHYUG-----DQVD--DEBDPLYDRLO		1041
Db	1104	VAAQROCVDAARLCQHGLCLVDAG--NTIHCROAGTGTGSEDLVDECSP-----	-----	1152
Qy	1042	DNGCGHPDASCADYLFDOTTVGVFHLRSPJGOYKILFDPAKCAKKAATATYANOLSYA		1101
Db	1153	---SPFCQNCATCTD-----LGGTSCKVAGYHGVNCSDEIDCLIS	-----	1190
Qy	1102	QKATYHILCSAGWLESGHVAYPPTYASOKCGANYGI-----	VDGSRANKSEM	1149
Db	1191	-----HPCONG--GTCLDLPNTYKC--SGPRGTQGHCEINVDCCNPVDPVSRSPK---	-----	1238
Qy	1150	MDVFCY-----RMKDVNCTKAGYVGVGSGFSGSGLLOVL		1183
Db	1239	----CFNNGTCVDVGVGYSCTCPRGVGE--RQGVVNBCL		1273

RESULT 4  
US-08-185-432-19  
; Sequence 19, Application US/08185432

```

1 Patent No. 5750652
2
3 GENERAL INFORMATION:
4 APPLICANT: Artavanis-Tsakonas, Spyridon
5 APPLICANT: Busseau, Isabelle
6 APPLICANT: Diederich, Robert J.
7 APPLICANT: Xu, Tian
8 APPLICANT: Matsuno, Kenji
9 TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
10 TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
11 NUMBER OF SEQUENCES: 23
12
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: PENNIE & EDMONDS
15 STREET: 1155 Avenue of the Americas
16 CITY: New York
17 STATE: New York
18 COUNTRY: U.S.A.
19 ZIP: 10036-2711
20
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: PatentIn Release #1.0, Version #1.30
26
27 CURRENT APPLICATION DATA:
28 APPLICATION NUMBER: US/08/185,432
29 FILING DATE: 21-JAN-1994
30 CLASSIFICATION: 530
31
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Mistrick, S. Leslie
34 REGISTRATION NUMBER: 18,872
35 REFERENCE/DOCKET NUMBER: 7326-006
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: (212) 790-9090
38 TELEFAX: (212) 869-8864/9741
39
40 TELEX: 66141 PENNIE
41
42 INFORMATION FOR SEQ ID NO: 19:
43 SEQUENCE CHARACTERISTICS:
44 LENGTH: 2703 amino acids
45 TYPE: amino acid
46 TOPOLOGY: unknown
47
48 MOLECULE TYPE: protein
49
50 US-08-185-432-19

```

```

Query Match          6.0%; Score 475; DB 1; Length 2703;
Best Local Similarity 23.1%; Pred. No. 7.3e-30;
Matches 268; Conservative 97; Mismatches 385; Indels 414; Gaps 75

OY      218 MCALGTASVWMPGVNGTGTCCGCGLGFNGTACE-----TCEGKYYGI--HCDQACS--- 264
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       359 ICVNWMAEL-DCSNNTDDCKQAACRYGATC IDGVGSFYCCQCTKRGKTGLLCHLDDACTSNP 417
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

OY      265 C-VHGRCSQGPL-GDGCDCDIDVNGRVKCDMEITTDNCN--GTCHTSANCLLDPDGKASC 320
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       418 CHADAICTSPITNGSYACSCATGYAGVDCSEI--DECDQGSPECHENGICVMTPL--GSYAC 474
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

OY      321 KCAAGFRNGTVC--TAINACETSNNGSGCSTKADCKRTTPGNRVYCVKAKAGTYG----DGIV 374
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       475 NCSQGF--TGPRCEININDCESH--PCQNEGSC-LDDPETFRVCMPGRTTGQCEIDIDE 529
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

OY      375 CLEINPCLNHHGCG-DR-----NAECTGTGP--NOAVCN----- 405
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       530 C-QSNPCL-NDOTCHDKI NGFKSCALGFTTGARCOINIDDQSOPCRNRGICHDISIAGYS 587
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

OY      406 --CLPKYTGDGKVCGL-INVCILT--NGGSPFAFCNTEDDQRCTKPKPYTGGIYVC 459
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       588 CECPPGTGTS--CEININDCDSNCPHRKC-----IDVNSFEKCLDDPGYT--GYIC 636
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

OY      460 RGSITGELPKNP-----STSOYFPOLOGHAHARELAGPGRFTYFALSSSFNHERPI 510
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       637 QKQI-NECESNPFQDFGHQDQRVSGSYCCQ-----AG-----TSCKNEVNV 678
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

OY      511 KDMDDQGLMSOVLRIHVHVCQDQLLDNLKVTTSATTLGQEPVSISSQDPVFIINNAKVL 570

```

Db 679 NE-----CHSNPC-----NNGAR- 691  
QY 571 SSDIITNGVIVHIDKLSPKNLI--TPKDALGRVLQNTTVAANHGYTKFSKLIDS 627  
Db 692 -----CIDGINSYKQCVPTGTQHCENKVECLSSPANNG----- 728  
QY 628 GLISVTDSHTPTVFWPTDKALEALPEEQDFLENQDNKDKLSYLKFIYIRDSKALA 687  
Db 729 ----VCIDQV-----NGYKCECPRGFYDAHCLSDVDECA 758  
QY 688 SDUPRSASMKTLQG--SELSTVRGIG-----SDIGELFLBQMKRFTHRGLIDVGVAY 739  
Db 759 SN-PCVNEGRCEDINIEFICHPRGYTGKRCCELLIDECSSNP--CQ--HGTCYDKLNAF 813  
QY 740 GIDCLLMPPLTGGRDPTFTFDIPGEC--GSCI-----FTPK-CPLKSPK 782  
Db 814 SCQC--MGTGYQKCEFTNIDCVINPCNGNGTCLDKVNGYKCVCKVPTGRCEKSKMPC 871  
QY 783 GVKR-----KCI--YNPLPR-----RNVEGQNL----- 805  
Db 872 ARNRCKNEAKCTPSSNFDSCCTCKLGYTGRYCDEDIDECSSLSPCRNGASCLANVPGSYR 931  
QY 806 CTVVIOPTRCCHGYFMPDC-----QACRGGPTPCNNRGMCDLITPMG--QCLCHTGPNG 859  
Db 932 CL-----CTKGYEGRCALINTDCC--ASFPCQNGRTCLD--GIDYSCICLDVDFDG 978  
QY 860 TACEL-----CMHGRF-----CPDCQPRSCSEHGQCD 886  
Db 979 KHCTEDINECLSPQCNATCSQYVNSTYCTCPLGFSINCQINDEDCTESSCLNGGSCI 1038  
QY 887 EGITGSGECLCETGWTAAACDPTPAVAVG--TPACSVHATCTENN--TCVCNLYEGDG 942  
Db 1039 DGINGY--NCSCLAGYSGANCQYK--LNKCDSNPLCATCHEQNNEXYTHCPSFTGK- 1093  
QY 943 ITCT--VUDEFCQNNNGCAKVAKCSQKQGVSCSKGKYGKGYSC-IEIDPCADGV----- 996  
Db 1094 -QCEYVDWCGQS--PCENGATCSOMKHQFSCKSAGWTGK--LCDVYTISCDADAARKG 1148  
QY 997 -----NGGCHFHATCRMTGPGKHKCEKSHVYGVGVDEPBDLPIDRCIDQNGOCHP 1048  
Db 1149 LSLRQLCNNGTKCKDYG-----NSHYCYSOGYAGS--YQCKE--IDEC--QSQPCQN 1194  
QY 1049 DASCADLYFDQTTGVPHLSPLQYKLTTPDKAKACAKAATATYVNLVYAAKAKYHL 1108  
Db 1195 GGTCRDL-----IGAYEQCQROGFGQNCENLIDDCANP----- 1229  
QY 1109 CSAWLESGRAYPTTASOKGANYGVITVYGRANKSEMDVFCY-----RMKDVN 1161  
Db 1230 CQNGGTCHDRV---MNFSCSCPPTMGIT---CEINNDCKPGACHNNGSCIDRVGGFE 1282  
QY 1162 CTCRAGYVGDGFGSGNLLQYL 1183  
Db 1283 CVCQPGFY--GARCEGDI NECL 1302

RESULT 5  
US-08-185-432-17  
; Sequence 17, Application US/08185432  
; Patent No. 5750652  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Buseau, Isabelle  
; APPLICANT: Diederich, Robert J.  
; APPLICANT: Xu, Tian  
; APPLICANT: Matsuno, Kenji  
; TITLE OF INVENTION: DETEX PROTEINS, NUCLEIC ACIDS, AND  
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York

COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELEPHONE: (212) 869-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2556 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-17

Query Match 6.0%; Score 471.5; DB 1; Length 2556;  
Best Local Similarity 20.6%; Pred. No. 1.3e-29;  
Matches 262; Conservative 113; Mismatches 427; Indels 471; Gaps 73;

QY 127 EKVLEIQKRNQDNNDITIVREGC--GKCSQAQAPLE-----TPKLAETR----- 169  
Db 256 ENIDDCPPNNCKNG-----GACVDGVNTVYNCPCPEPTGQYCTEDVDECOIMPAQONG 309  
QY 170 -----KCIYSIYPMGRKSVFTGCOPOCVRTI--ITRACWLASIAHNAK----- 210  
Db 310 GTCHNTFHGGVNCVCMNGTGE-----DQSENIIDOCASACHGATCHDRVASFCE 360  
QY 211 -PARGEVKCALGTASV-----WDGVNGTGTQCGLGIFNGTAC-----ETCTEKG 254  
Db 361 CPHGRTGLLCHLNDACISNPNEGSNCDDTNPVNGKAICTCPSGYGPACSDVDVDECSLG- 419  
QY 255 YGIHDDQACSVHNGCSGRLPDGSCDDVDGVRGKCMETITTMNC-NGTCHTSAKCLLD 313  
Db 420 -ANPEHAGKCTIN-----TLGSEFCQCLQGYTGPRC--EIDVNECVSNPCQNDATC-LD 469  
QY 314 PDGKASCCAGAFRNGNTVCTAINA--CETS-----NGCSTKADCKRTTPGNRVCSKAG 367  
Db 470 QIGEQCMCMRGYES---VHCEVNTDEGCASSPCLHNGKCLDKIN-----EFGCECTPG 519  
QY 368 YTGDIIVC-LEINPCLNHHGCDRNAECTGTGPNQAVNCCLPKYTGDKVCSLINVCLTN 426  
Db 520 FTGH--LCQYDVDEGAST--PCKNGAKCLD-GPNITYTCVCTEGYTG-----THCEVD 566  
QY 427 NGGCSPFACWYTEDQDL-----CTCKPDYTG-----DGLVCGSITYGELPKRPS 472  
Db 567 IDECDPDP-CHYGCKDKDVAFTPTCLCRPGYGHHCETNINECCSQPCR--LWGTG-QDPD 622  
QY 473 TSOYFQLOEHAHVREIAPGPFYVAPLSSSFNHEPRITKMDQQLMSQVLRHYHVGQO 532  
Db 623 NAVLCE-----CLKQTTGP-----NCEINLDD-----CAS 647  
QY 533 LLDNLKVYTSATTLQGPVSIYSQDTVFJNNEAKVLSDDIISTNGVIVHIDKLSPKN 592  
Db 648 SPCDS--GTCLDKIDGYEACCEPGYTGSMNSNIDFCAGNPHNGG----- 691  
QY 593 LLITPKDALGRVLQNTTVAANHGYTKFSKLIDSGLISVTDSHTPTVFWPTDKALE 652  
Db 692 ----TCEDEGI-----NGFYCRCEGY-----HDPFTCLSEVNECSNDCV----- 726  
QY 653 ALPPEQDFLENQDNKDKLSYLKFIYIRDSKALASDLPRSASMKTLGSGELSVRC----- 708

Db 727 -----HGACWMSLNG-----YCCDDP 743  
Qy 709 ---GTGSDIGELFLNEQMC-----RFTHR 729  
Db 744 GWSGNCINNNECSBNSNGCTCKDMTSGIVCTWEGFSGPNCOTINIENCSNPCLK 803  
Qy 730 GLEFDVAVAGIDCLLMPRTLGRCDDFTTFDIPGEC---GSC-----JFPKPLKS 779  
Db 804 GTCIDVAVAGYCNCLL---PYTGATCEVVLAPCAPSPCRNGECRSEDYESTSCVP--- 858  
Qy 780 KPKGVK-----KKCIYNPLPFRNVEGCONLCTVYIOTPRCHGYFMPDQCAPGP 831  
Db 859 -TAGAKGCTCEVDINECVLSPC---WHGASCON-----THGXRYCHQAGYSGR 903  
Qy 832 D-----TPCNRBMCRD-LYTPMGOCCLHTGFNGTACE----- 863  
Db 904 NCFDIDDCWPNPCNNGSCTDGINJAF---CDCLPFGWGTCEEDINECASDPCPNANG 961  
Qy 864 -----LCMHGRFG-----PDCQPRSCSEHGOCDEGITSGECLCETGWTAAASCD 907  
Db 962 TDCVDSYTCCTCPAGFSGIHCEHNTPPDCTESSCFNGTCYVDGI-NSFTCLCPGFTGISTCQ 1020  
Qy 908 TPTAFVAVC-TPACSVHATCTEN---NTVCNLYNEGDIITC-TVVDFCKONNGSCAVYA 962  
Db 1021 ---HVNVNCDRPPCLLGGTCQDGRGLHRCCTPOGYTGP---NQONLVHMC---DSSPCXKNG 1073  
Qy 963 KCSQKGTQVSSCKKGYG-----DGYSCIEFDPCADVNGS---CHEHATCMTPGKHK 1015  
Db 1074 KCMOHTHOYRCBPBGWGLYCDVPSVSC-EVAAQROGVDAVARLOHGGLCVDA-NTNH 1131  
Qy 1016 CEKSHYVG---DQVD-CEPEOLPLDRGLDNGQCHPDASCADLYFODTIVGVHLSRP 1070  
Db 1132 CRCQAGYGVSGEDLYDCCSP-----SPCQNGATCTD-----LGGSCKCV 1173  
Qy 1071 LGQYKLTFDKAKKACAAATATATYNNLSYAKAKYHLCASGWLSEGVANPTTYASOKC 1130  
Db 1174 AGYHGVNCSSEIDELCS-----HPCQNG---GTCLDLPNTY---KC 1208  
Qy 1131 GANVGIYDYSBRANKSEM-----MDVFCY-----RMADVNCTGACVYG 1170  
Db 1209 SC-----PMGTQVHCINVDCCNPVDPVSWSPKCFNNTCVDVQGYSCYCPGPHVG 1262  
Qy 1171 DGFSCSGLLOYL 1183  
Db 1263 E--RCEGDVNECL 1273

RESULT 6  
US-08-185-432-16  
Sequence 16, Application US/08185432  
Patent No. 5750652  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Busseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
NUMBER OF SEQUENCES: 23  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-16

Query Match 5.7%; Score 450.5; DB 1; Length 2471;  
Best Local Similarity 22.7%; Pred. No. 6.9e-28;  
Matches 270; Conservative 108; Mismatches 421; Indels 391; Gaps 73;

Qy 133 OKNRCDNDITIIIVREGCKC-----SQAPCPLERKPL-----RET 168  
Db 69 EKNRQNGCTVYAOAMLKATRCASGFTGEDCOYSTHPCFVSPPCLNGCTHMLSDT 128  
Qy 169 KCIYSIYFMGRKRSYVIGCQPOCVRTIITRACMLASLANMKPAPGEVYKCALGASVMD 228  
Db 129 YECQVQVFTGKE-----CQ-----WTDAC---LSH-----PCANG---STCT 160  
Qy 229 GYNGTGCQCGGFGNGTACET-----CTGKGVIIHDDQ 261  
Db 161 TVANFSCKCLTGFYQGCETFDVNECDIPGHQHGCTLNLPGSYQCCPGGFTGYQYDS 220  
Qy 262 ---ACS---CVH-GRCSGPLDGS---CDDVGMRGVACDMEITTDNC-NGTCHTSANCL 311  
Db 221 LYPCAPSPCVNNGICRQ---TGDFTEECNCLPGEFGSTCERNI---DDCPNHRQNGVYCV 276  
Qy 312 LDPDGKASCACAGFRNGVYCTA--INACETSNNGCSTKADCKRTTPGNRYVCVCKAGYTG 370  
Db 277 ---DGVNTYNGRCRPPOMTGOCTEDVDECLLDPPNACQNGGTCANRN-GGYGCVYVNGSG 332  
Qy 371 DGIYCLEINPCLENHGGCDRNAECTGTG---PNOAVNCCLPKYTGSGVCSLINVCLNN 427  
Db 333 D-----DSENIIDC-ARASCTPGSTCIDRVASFSCMCEKAGAGLCHLDACISN- 382  
Qy 428 GGCSPFAFCNYTE--ODQRICTCKPDYTG-----DGIYCRGSIYGLPKNPSTSQYFFQ 479  
Db 383 -PCHKGALCDINPLNGYICICPOGYKCADCTEDVDECBAMNS-----NPC----- 427  
Qy 480 LOENAVRELAGRPFTVAPLSSSFNHEPRIKDWDOGLMSOVLRYHVYVCOQ---LLLDN 537  
Db 428 --EHAGKCVNTDGA---HCECLKGYAGRPCENDINCHSDPQONDATCDHK 474  
Qy 538 LKVTTSATF---LQGEPRVSIYS--ODYVFNNKAYVLSSDIS-----TNGVYHV 583  
Db 475 IGGFTCLCMGPFGYHCELEINECOSNPNVNGQCV---DAVNRFOCLCPGFTGPVQOI 531  
Qy 584 -IDKLSPKNLLITPKDALGRVLONLTVVAAHGY-----TKFSKLIODSGLSVITDSI 637  
Db 532 DIDCCS-----TP-----CLNGAKKIDHPNGEGCCATGFTGYLCEENINDNDPPPC 579  
Qy 638 HTPVTFWPTDKALEALPPEODPLFNQDNKDKLSYLKFIYIRDSKALASDLPASAWK 697  
Db 580 H-----HGOCODGIDSYCTICINPGYMAICSD----- 606  
Qy 698 TLQGSSELSVRGCTGSDIGELF-----LNEQKGRFTHRGLFDVGVAVYGDCLLMNPTLGR 753  
Db 607 -----QIDECYSSPCLNDGRC-----IDLNVGYQCNQPGTSGVN 641

```

0Y 754 CDTFTTDD-----IGEC-----GSCIFPP-----KCLPSKPKYKKKCIYNLPF 795
Db 642 CE--INFDDCASNPCIHGICMDGINRSCSCSGYGTGORCNDI-----DDCASNp-- 690
0Y 796 RRNEGCNCLCTVYIOTPRCCCHGYFMFDCCACPGPDTPCNNRMCRCMDLYTEPMGCL-- 852
Db 691 -----CRKAT-----CINCVNFRRC-ICEGRHHP-----SCISQVNBCLSLN 728
0Y 853 -----CHTGNGTACELCHMGRFGRPCQ-----PRSCSEHQCDDEGITSGECLCETG 900
Db 729 CIGHNCTGGSLGYAC-ICDAGWGWINEVDKNECLSNPCONGTCTCNLVNGY-RCCTCKKG 786
0Y 901 WTAASCDPTAFVAVCTPACSVHATCTEN---WTVCYNLNEBDDGITC-TVVDPFCKONG 956
Db 787 FKGINCO--VAIIDCASNPCINOGTCTDDLSGTTGCHCLPYTK--NCTOTVLAPCSBNPC 842
0Y 957 GCAYAVAKCSOKGTGVSCSKKGYKGDGYSC-IEIDPCADGVNCGCHEATCTMTPGKHK 1015
Db 843 ENAAVCKESPFESEYETCLCAPMGWGO--RCTIDIDEC--ISKPCMHGLCHNT--QGSYM 896
0Y 1016 CEKASHVCGDCEDEPRLDLCRLDONGGCHDASCADLYRQDTPYGVGFHLRSPLOYK 1075
Db 897 CECPGPF--SCMDCEED--IDDLPLAN--PCONGGSCMD-----GVNTF--SCLCLPG 940
0Y 1076 LTFPRKAK-----EACAKEAATIAVYNLSTYAKAKYHLCSAGULESGRAVAPPTYAS 1127
Db 941 FTGDKCQIDMNECLSEPC-KNGGCSBYVA-SYTKC-----QQAGS-----EDG 981
0Y 1128 QKCGANVVGIVDYSRANKSEMDVFCYR-----MKDVNCTCKAGYVG 1170
Db 982 VHCENNI-----NECTESSGFNGTGVDSINSESLCLCPVGTG 1019

```

```

1      RESULT 7
2      US-08-083-590A-19
3      ; Sequence 19, Application US/08083590A
4      ; Patent No. 5786158
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: Artavanis-Tsakonas, S. et al.
9      ;
10     ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
11     ;
12     ; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
13     ;
14     ; TITLE OF INVENTION: Nucleic Acids
15     ;
16     ; NUMBER OF SEQUENCES: 21
17     ;
18     ; CORRESPONDENCE ADDRESS:
19     ;
20     ; ADDRESSEE: Pennie & Edmonds
21     ;
22     ; STREET: 1155 Avenue of the Americas
23     ;
24     ; CITY: New York
25     ;
26     ; STATE: New York
27     ;
28     ; COUNTRY: U S.A.
29     ;
30     ; ZIP: 10036
31     ;
32     ; COMPUTER READABLE FORM:
33     ;
34     ; MEDIUM TYPE: Floppy disk
35     ;
36     ; COMPUTER: IBM PC compatible
37     ;
38     ; OPERATING SYSTEM: PC-DOS/MS-DOS
39     ;
40     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
41     ;
42     ; CURRENT APPLICATION DATA:
43     ;
44     ; APPLICATION NUMBER: US/08/083,590A
45     ;
46     ; FILING DATE: 25-JUN-1993
47     ;
48     ; CLASSIFICATION: A35
49     ;
50     ; ATTORNEY/AGENT INFORMATION:
51     ;
52     ; NAME: Mistrock, S. Leslie
53     ;
54     ; REGISTRATION NUMBER: 18,872
55     ;
56     ; REFERENCE/DOCKET NUMBER: 7336-015
57     ;
58     ; TELECOMMUNICATION INFORMATION:
59     ;
60     ; TELEPHONE: 212 790-9090
61     ;
62     ; TELEFAX: 212 8698864/9741
63     ;
64     ; TELEX: 66141 PENNIE
65     ;
66     ; INFORMATION FOR SEQ ID NO: 19:
67     ;
68     ; SEQUENCE CHARACTERISTICS:
69     ;
70     ; LENGTH: 2471 amino acids
71     ;
72     ; TYPE: amino acid
73     ;
74     ; STRANDEDNESS: single
75     ;
76     ; TOPOLOGY: unknown
77     ;

```

MOLECULE TYPE: peptide  
US-08-083-590A-19

Query Match 5.7%; Score 450.5; DB 1; Length 2471;  
Best Local Similarity 22.7%; Pred. No. 6.9e-28;  
Matches 270; Conservative 108; Mismatches 421; Indels 391; Gaps 73;

[illegible]

Db 843 ENAAVCKESPNESTYTCACAPGMOGO--RCTIDIDEC---ISKPCMNHGLCHNT--QGSYM 896  
QY 1016 CECKSHYVGDVDCPEBOLPLDRLQDNGQCHPDASCADLYQDPTVGFHLRSPUGYK 1075  
Db 897 CECPPGF--SGMDCEED---IDDCLAN--PCONGSGCMD-----GVNPF---SCLCLPG 940  
QY 1076 LTFDKAK-----EACAKEATATATYNOLSYAOKAKYHLCAGWLESGRVAYPTTAS 1127  
Db 941 FTGDKCQDTMNECLSEPC-KNGGTCSDYVN-SYTKC-----COAG-----FDG 981  
QY 1128 QKCGANVGIYDYSRANKSEMMDVFCYR-----MKDVNCTCKAGYVG 1170  
Db 982 VHCENNI-----NECTESSCFNGGTCVDGINSFSCLCVPVFTG 1019

RESULT 8  
US-08-532-384-19  
Sequence 19, Application US/08532384  
Patent No. 6083904  
GENERAL INFORMATION:  
APPLICANT: Artavaonis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,384  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/083,590  
FILING DATE: 25-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-532-384-19

Query Match 5.7%; Score 450.5; DB 3; Length 2471;  
Best Local Similarity 22.7%; Pred. No. 6.9e-28;  
Matches 270; Conservative 108; Mismatches 421; Indels 391; Gaps 73;

QY 133 QKRNCDNDNTIIVRGECKC-----SQAPCELETKPL-----RET 168  
Db 69 ENNRONGSTCAVQAMLKATCRCAAGTGEDCQYSTHPCFVSRCPLNGTCHMLSRPT 128  
QY 169 KCIYSIYFMGRSVFIGCQPCQVFTIITRACWLASLAHNAKPADGEVWKALGTAQVMD 228

Db 129 YECTQVGFTEKE-----CQ-----WTDAC-----LSH-----PCANG---STCT 160  
QY 229 GVNGTGTCCQCGIFNGIACET-----CTESKYIHCDQ 261  
Db 161 TVANPFSCKCLGTGFKCETIDVNECDIPGHQHGSTGLNPGSYQCCPQGFSTGYCDS 220  
QY 262 ---ACS---CVH--GRCSQGLADGS--CDGDVGMKGVKCDMEITTDNC--NGTCHTSANCL 311  
Db 221 LTVPCAPSPCVNGVGTGRQ--TGDFTEPCNCLPGEFSGTCEIRNI--DDCPNHRQMGNGVCV 276  
QY 312 LDPDGKASCKAAGFRGNQTVCTA--INACETSNGGCSTKADCKRTTPGRVRCVCAGYTG 370  
Db 277 --DGVNITNCKCPQWMTGQFCTEDVDECLQPNACQNGTCAANN--GGYGCVCVNGSWG 332  
QY 371 DGIYCLEINPCLENHGGCDRNAECTQTG--PNAVCNCLPYRTDGRKCSLINCLTNN 427  
Db 333 D-----DCESENIDDC--AFASCTPGSTCIDRVASFSCMEPKGAAGLLCHLDDACISN- 382  
QY 428 GGCSPFAPCNYTE--ODRITCTCKPDYTG-----DGIYCRGSIYELKPNSTQYFPO 479  
Db 383 -PCHKGALCDTNPNGOYICTCPQCYKGGADCTEDYDECMANNS-----NFC----- 427  
QY 480 LOEHAVRELAGPPTVTFAPRLSSFNHPRIKMDQOQGLMSQVLRHYVCGQO--LLLDN 537  
Db 428 --EHAQKCVNTDGAH-----HCECLKGVAQPCEMDINCHSDPCOMDATCLDK 474  
QY 538 LKVTTSATY--LOGEPVSIYS--ODTVEINNEAKVLSDDIS-----TNGVIHV 583  
Db 475 IGGFTCLCMKPGFKGVHCELINEQSNPCVNNQCV--DKVNRQCLCPGFTGVQCI 531  
QY 584 -IDKLISPKNLLITRKDALGRVLNITVAANHGY-----TFESKLIJDSGLSYITBSI 637  
Db 532 DIDDCSS-----TP-----CLNGAKCIDHPNGYEQCATGTGVLCEENIDNDPDPG 579  
QY 638 HTPVTVFMTKALEALPEEQDQDFLENDKDKLSYLKFHVIRDSKALASDPRSASMK 697  
Db 580 H-----HGOCQGDISTYITCIPGTGATCSD----- 606  
QY 698 TLQSELSVRCGTGSDIGELF--LNEQMCRFIRHGLFDVAVAYGIDCLLMNPLTGR 753  
Db 607 -----QIDBCYSSPLNDGRG-----IDLNVGYCNCQPGTSGVN 641  
QY 754 CDTFTTFD-----IPGEC-----GSCITFP-----KPLKSKPKGVKKKCIYNPLPF 795  
Db 642 CE--INFDCASNPCIHGCMGIDNRYSCVSPGFTGRCNIDI-----DDCASNP-- 690  
QY 796 RNVGECQNLCTVVLQTPRCCHGYFMPDQACPGGDPFCNNRGMCRDLTYTPMGOCL-- 852  
Db 691 -----CRKGAT-----CINGVNGFRC--TCPEGFHP-----SCYSOVNECLSNP 728  
QY 853 -----CHTGFNGTAGELCMHGRFRGPDQ-----PRSCSEHQQCDDEGITYSGECLCETG 900  
Db 729 CIGHNGTGLSGYKC--LDCAGWVGINCEYDKNECLSNPCQNGSTGDNLVNGY--RCTCKRG 786  
QY 901 WTAASCDTPTAVFAVCTPRACSVHATCTEN--NTCYCNLNTAGDGTTC--TYVDFEKQNNG 956  
Db 787 EKGYNQ--VNIDECASNPCLNQGTCEFDISGYTCHVLPYTK--NCOQTVLAPSPBNC 842  
QY 957 GCATYAKCSQKGTQVYSCSKKGYKGDGYSC--LEIDPCADGVNGGCHENATCMTGSGKHK 1015  
Db 843 ENAAVCKESPNESTYTCACAPGMOGO--RCTIDIDEC---ISKPCMNHGLCHNT--QGSYM 896  
QY 1016 CECKSHYVGDVDCPEBOLPLDRLQDNGQCHPDASCADLYQDPTVGFHLRSPUGYK 1075  
Db 897 CECPPGF--SGMDCEED---IDDCLAN--PCONGSGCMD-----GVNPF---SCLCLPG 940  
QY 1076 LTFDKAK-----EACAKEATATATYNOLSYAOKAKYHLCAGWLESGRVAYPTTAS 1127  
Db 941 FTGDKCQDTMNECLSEPC-KNGGTCSDYVN-SYTKC-----COAG-----FDG 981  
QY 1128 QKCGANVGIYDYSRANKSEMMDVFCYR-----MKDVNCTCKAGYVG 1170  
Db 982 VHCENNI-----NECTESSCFNGGTCVDGINSFSCLCVPVFTG 1019



```

1      RESULT          9
2      US-08-400-159-2
3      ; Sequence 2, Application US/08400159
4      ; Patent No. 5869282
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Ish-Horowitz, David
7      ; APPLICANT: Henrique, Domingos M.P.
8      ; APPLICANT: Lewis, Julian H.
9      ; APPLICANT:
10     ; APPLICANT: Myatt, Anna M.
11     ; APPLICANT: Fleming, Robert J.
12     ; APPLICANT: Artavanis-Tsakonas, Spyridon
13     ; APPLICANT: Mann, Robert S.
14     ; APPLICANT: Gray, Grace E.
15     ; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
16     ; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
17     ; NUMBER OF SEQUENCES: 20
18     ; CORRESPONDENCE ADDRESS:
19     ; ADDRESSEE: Pennie & Edmonds
20     ; STREET: 1155 Avenue of the Americas
21     ; CITY: New York
22     ; STATE: New York
23     ; COUNTRY: USA
24     ; ZIP: 10036-2711
25     ; COMPUTER READABLE FORM:
26     ; MEDIUM TYPE: Floppy disk
27     ; COMPUTER: IBM PC compatible
28     ; OPERATING SYSTEM: PC-DOS/MS-DOS
29     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
30     ; CURRENT APPLICATION DATA:
31     ; APPLICATION NUMBER: US/08/400.159
32     ; FILING DATE: 07-MAR-1995
33     ; CLASSIFICATION: 435
34     ; ATTORNEY/AGENT INFORMATION:
35     ; NAME: Mirock, S. Leslie
36     ; REGISTRATION NUMBER: 18,872
37     ; REFERENCE/DOCKET NUMBER: 7326-029
38     ; TELECOMMUNICATION INFORMATION:
39     ; TELEPHONE: (212) 790-9090
40     ; TELEFAX: (212) 869-9741/8864
41     ; TELEX: 66141 PENNIE
42     ; INFORMATION FOR SEQ ID NO: 2:
43     ; SEQUENCE CHARACTERISTICS:
44     ; LENGTH: 1404 amino acids
45     ; TYPE: amino acid
46     ; TOPOLOGY: linear
47     ; MOLECULE TYPE: protein
48     ; US-08-400-159-2

```

```

Query Match          5 3%: Score 413.5; DB 2; Length 1404;
Best Local Similarity 19.7%: Pred. No. 3,4e-25;
Matches 274; Conservative 127; Mismatches 482; Indels 507; Gaps 66;

QY  QCGLGFNGTACET-----GTEGKYGIHCDOA-----CSCVHGRC 270
    || : : ||
Db  QCAVYIYNTTCTTFCRPRBDPGHYACGSEGGKLLNGMGVNCBEALCKAGDPYHGRC 295
    || : : || : : || : : || : : || : : || : : || : : ||

QY  271  SGGPLGDSGCDGVRGKVCMDMETTTDNC -NGTCHTSANCLLDPDGKASCCKAAGFRGN 329
    || : : || : : || : : || : : || : : || : : || : : || : : ||

Db  296  DR-----PECEBERPWRMGRLCNECWVYFGCKHGCSCNGSA-----WRCVCDTIMG 340
    || : : || : : || : : || : : || : : || : : || : : || : : ||

QY  330  GIVYCA -INACTSNGGCTKADCKRTTPGNRVYCYCKAKYTDGIVCLEINCLENHGGC 388
    || : : || : : || : : || : : || : : || : : || : : || : : ||

Db  341  GILCQDILFEGST -HEPCKHGCTCENTAPDKYRCYCAEGLSGEGEIVE -HPCATR -PC 396
    || : : || : : || : : || : : || : : || : : || : : || : : ||

QY  389  DNNAECTGTGPNQA-----
    || : : || : : || : : || : : || : : || : : || : : || : : ||

Db  397  RNGGCTLTKTSRRTQAQYRTSHGRSNMCRPYBRSSMSLDHLRPEGALNGSSSGLY 456
    || : : || : : || : : || : : || : : || : : || : : || : : ||

QY  403  -----VCNCLPRYTGDKGVCYLINVCLLTNNGGCSPPAFCCNTEEDORICT 447
    || : : || : : || : : || : : || : : || : : || : : || : : ||

```

Dd	457	SLSLQLOLOQADLPPTCDCAAGT--GPICE-INIDECAGGCGEIGTCIDLDIGFR-CE	512
Oy	448	CKPDYTGDIYVCRGSIYGLPKNPSTSOYFPOLOEHAVELAGPGPFTVFAFLSSSFNHE	507
Dd	513	CPPEMHGD--VCQVD-----VNECBAP-----	532
Oy	508	PRKIDMDQGLMSOVLRYHVYGCQOULLDNLKAYTTSATTLOGEPVSIYSQDYFINNEA	567
Dd	533	-----HSAG-----IAANALLTPTATVALIISNLS-----STALLAALT	565
Oy	568	KVLSSDIISTNGYIHWIDKLILSPKNLLITPKALGV--LOMLTYVA--ANHGYSFKL	623
Dd	566	SAVASTSLAIGPCINAKECNPQSPACIOKEBGGVTCAEMLDVCQGCRRGATCID-L	624
Oy	624	IQD-----SGLSLVITDSIHPTVPTFMPETDALKALELPEQODPLFNODNKDKLSY LKF	677
Dd	625	VNDRCACASGFTG-----	638
Oy	678	HVIRDSKALASDLPRASAMKTLQGSLSYRCGTSGDIGELFLNEQMCRFHIGLLFDYGV	737
Dd	639	--RDCE--TDIDECATSPCRNGE-----C-----YDM	662
Oy	738	AYGIDCLLAMPITLGRDPTTTTDFDIGEC--GSCITFR-----CPLSKPRGKKKICY	790
Dd	663	VGKEKICICPLGYSLSLCEAEKENCSTPCLEGHCLNTPBGYYCHCP--PDRAGHC--	716
Oy	791	NPLPFRNRVBECONLCTVAVIOTPRCHGF--	830
Dd	717	-----EQLRPLCS--QPPCEGCFANVSLATSTTTTTTTTTTATTTRKRAKNSG	763
Oy	831	PDTPCNKRGMC-RDLYTPMGQCLHTGTNGTACELCWHGREGPDCQPRSPSEHGQCEG	888
Dd	764	--LPCSHGSGCEMSDVGT--FCCHVGHGTGFC--HNL--NECSPKCHNGICIDG	813
Oy	889	ITGSGECLCEFGWTAASCDTPAVFAVCPACSVNATCTENNTVCYNLNEBDGTTCTYV	948
Dd	814	DGFTGCEGSGMWG-----KCSERTQCYAGQC-----	842
Oy	949	DPCRONNGCAKAYAKCSOKGTQYSCSKRGYKQDGYSCIE-IDPCADGVNGCHEHATCR	10070
Dd	843	-----QNGTCTMPCA--PDKALQPHCHCAQGM--TGIFCAEALDQCR--GQPCHHGTCSE	891
Oy	1008	MTGPGKHRCCECKSHYVGDVDCEREDLPDRCLQDNGQCHPDASCADLYFDPTTYGVFHL	1067
Dd	892	SGAGMFRVCQAQGF--SGPDC--RLNNECSPO--PCQGATCID-----GIGYSC	937
Oy	1068	RSLPGOT-----KLTFDKAKACAKKEAATYATYNOJLSAOKAYHILCSAGULE--SGVA	1120
Dd	938	ICPGRHGLRCEILIDSPKSCACONASNTISPYTALNRSQ-----NMLDIALTGR--	986
Oy	1121	YPTTYASOKGCANVGVIVDGSRANKSEMDWVCYMKKOVNCTCKAGYVGGQFSGSGLL	1180
Dd	987	---TEDENCA--CVCENGTSRCTNLM-----CGLFNCTKVDLSLSSMLS	1028
Oy	1181	QVLASFSLTNFLETVLAFESKSSARQAFILKHLTDLSTIGTLFVQNSGLPGKNSL---	1236
Dd	1029	GVCKQHEVCVPALSEFCLSSPCVWRDC--RALEPSRRVAPRILPAKSSCMQNPQAVVEN	1086
Oy	1237	-----SGDIEHLTNVVSFYNDLVN--GTF-----LRTMLG	1267
Dd	1087	CARLTILALERVGKASVEGLCSILVRVILIAOLIKKPAFSGODPGMVLICDICTGTN	1146
Oy	1268	SOLLITFSOOLHO-----ERRPYDG--SLILOMDILIANGLIHISPL	1310
Dd	1147	DTVELTYSSSKLNDPOLPAVAVGLLGELLSSRQNGIQRRKELELOHAKIALATLSIVEYKL	1206
Oy	1311	RAPPTAATAHSGLTGIFCAV--VVTATAIALAAYSYRLKQRT-----	1354
Dd	1207	ETAVAVAGSGSGS--LLIGVLCGVFIYLVGFSVFSISLWKQRLAYRYSNGMNLTPSLDALRH	1265
Oy	1355	-----GRFDPDKRTIMSKULASSPRLISOT--LCMRPQRHQPSPVPTPSOTLENRI	1405
Dd	1266	EEKSSNNLQNDENIRRTYTNLKGSTSLAAGOMELSLNP-----APELAAASAASSAL	1319

QY 1406 WRTATLMGHC 1415  
 Db 1320 HRSQPLFPCC 1329

RESULT 10  
 US-08-611-729A-2

; Sequence 2, Application US/08611729A  
 ; Patent No. 6004924

; GENERAL INFORMATION:

; APPLICANT: Ish-Horowitz, David  
 ; APPLICANT: Henriquez, Domingos M.P.  
 ; APPLICANT: Lewis, Julian H.  
 ; APPLICANT: Myatt, Anna M.  
 ; APPLICANT: Fleming, Robert J.  
 ; APPLICANT: Aravanis-Tsakonas, Spyridon  
 ; APPLICANT: Mann, Robert S.  
 ; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
 ; TITLE OF INVENTION: SERATE GENE AND METHODS BASED THEREON  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/611,729A  
 ; FILING DATE: 06-MAR-1996  
 ; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mirock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 7326-037  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864

; TELEFAX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1404 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; US-08-611-729A-2

Query Match

Best Local Similarity 5.3%; Score 413.5; DB 3; Length 1404;  
 Matches 274; Conservative 127; Mismatches 482; Indels 507; Gaps 66;

QY 237 OCGLGNGTACET-----CTEGYTGTHCDA-----CSCYHGKC 270  
 Db 236 OCAVYTYNTCTTCFPRRDQFGHYACGSEGOKLCLNGWQGVNCEBAICKAGCDPVHGKC 295  
 QY 271 SOGPIGDSGCDQVGRGVKCDMEITDNC-NGTCHTSANCLLDPPGKASKCAAFGRGN 329  
 Db 296 DR-----PGCEBCEPGRGRLCNECMYRPGCKHSGNSA-----WKVCVCTNMG 340  
 QY 330 GYVCTA-INACETSNCGSKTADCKRTTPGNRYVCVCAKAGYTGDIIVCLINPCLNHHGCG 388  
 Db 341 GILCDQDLNFCGT-HEPCKHGTCENTAPDKYRCCTACGLSGGCEIVE-HPCATR--PC 396  
 QY 389 DRNAECTGTGPNQA-----402

Db 397 RKGCTTLKTSNRTQAQYRTSHGRSNMGRPYRBRSSSMKSLDLHLPREGALNGSSSGLY 456  
 QY 403 -----VCNCLPKRYTGDKVCSLINVCLINNGCSPFAFCNTYEDQRICT 447  
 Db 457 SLGSLQLOOQALPDPFTCDCAAGMT--GPICE--INIDECAGCGEHCRTIDILIGFR--CE 512  
 QY 448 CKPDYTGDCIVCRGSLYGLPKNPSTSOYFPQLOEHAVALAGPGRFYFAFLSSSFNHE 507  
 Db 513 CPPEWHGD--VQVD-----VNECBAP-----532  
 QY 508 PRKIMDQGLMSOYLRYHVCCOOLLNLTATTSATLQGEPIVSIYSODTFINNEA 567  
 Db 533 -----HSAG-----IAAMALTTATATISNLS-----STALLALT 565  
 QY 568 KYLSDDIISTNGVIHYDKLSPKMLLTPKDALGRV--LQNLITVA--ANHGRTFSKL 623  
 Db 566 SAVASTSLAIGPICINAKECRNQPGSFACICKGCGVGTCAENLDPCVGCRCRATCID-L 624  
 QY 624 IOD-----SGLLSYTTDSIHPVTVFPTDALEALPPEQDLEFNQDKKLKSLYKF 677  
 Db 625 VNDYRCACASGFTG-----638  
 QY 678 HVIKSKALASDLPRASMKTLQGSLSVRCGSDIGELFNEQWCRFIHRLLEFVGV 737  
 Db 639 ---RDE--TIDECATSPCKNGE-----C-----VDM 662  
 QY 738 AVGIDCLMNPPLGRCRDPFTFPDIPGEC--GSCJFTPK-----CPKSKPGVKKKCIY 790  
 Db 663 VKFNCICPLGYSGSLCEAKENCTSPCLTEGHLMTPEGYCHP-----PBRAGHC-- 716  
 QY 791 NLPFRNVEGCONICTVVIQTPRCCHGF-----MDCQACPGG 830  
 Db 717 -----EOLRPLCS--QPPCEGCFANVSLATSATTTTTTTTATTRKMAKPSG 763  
 QY 831 PDPCCNRMGR--DLTPMGOLCHTGFNGTACELCWMGRGRCPPDQPCSEHGGCDDG 888  
 Db 764 --LPCSGHSCSEMSDGT--FCKVHGTTFCE--HNL--NECSPPCRNGGICLDG 813  
 QY 889 ITSGECLCETGWTAAACDTPAVFAVCPACSVHATCTENNTVCNLYEGDGLTCTVY 948  
 Db 814 -DGFTECEMSGWTG-----KRCSEKATGCVAGG-----842  
 QY 949 DPCKONNGCAVYAKCSOKGTGVSCSKKGYGDIYSCLIE-IDPCADGVNCGCHEATCR 1007  
 Db 843 ---QNGTCMGA--PDKALQPHRCAPGW--TGFECAEALDOCA--GODPHNGTGE 891  
 QY 1008 MNGPGKHCKECKSHYGVGVDEPDLPLDRLQDNGGCHPAPASCDLYEFTTVGVFHL 1067  
 Db 892 -SGAGMFRVCVCAQGF--SGPDC--RINVNECSPO--POGGATCID-----GIGGYSC 937  
 QY 1068 RSPLEGY---KLPDKAKKAEKAEATATATYNOISYAGKAKYHLCAGMLE--SGRYA 1120  
 Db 938 ICPRGHGLRCELISDPRASQONASNTISPYTALNRSQ-----NMLDIALTRG-- 986  
 QY 1121 YPTTASOKCGANVGIYDYGSRANKSEMDVFCYMKKDVNCTCKAGYVGDFSGGSL 1180  
 Db 987 ---TEDDEMCNA---CVCENGTSRCTNLM-----GGLPCYKVVDPILSKSNLS 1028  
 QY 1181 QVLMSPSLNPLFLEVLARSSKARGQAFKLKHLTDLSTIRGTFLEVPONGSLPGKSL---- 1236  
 Db 1029 GYCKQHEVCVPALSETCLSSPCNVKDC--RALEBRRAVAPPLPAKSSCMQAVNVNEN 1086  
 QY 1237 -----SGRDIEHLITNVVSFYNDLVN--GTF-----LRTMLG 1267  
 Db 1087 CARLITLLALERKGAASVGLCSLVRVLLAQLIKKPASTFEGQDGMVLCDLKTGTN 1146  
 QY 1268 SOLITFSODOLHO-----ETRYVDG--RSIIQMDIIAANGIIHITSBL 1310  
 Db 1147 DIVELTVSSSKLNDPOLPAVAVGLAGELSSROLNGIQRKKELELOHAKLAALTSIVEVL 1206  
 QY 1311 RAPTATATAHSLGTGICAV--VLVTGAIALAAVSYFLKORT-----1354  
 Db 1207 ETARVADSGSHS-LIIGVLCQVFIYLVGFSVFLISLYWKORLAVRTSSGMNLPISDLAKH 1265





```

1 REGISTRATION NUMBER: 31,815
2 REFERENCE/DOCKET NUMBER: P-UW 2637
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (619) 535-9001
5 TELEFAX: (619) 535-8949
6 INFORMATION FOR SEQ. ID NO: 7:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 1010 amino acids
9 TYPE: amino acid
10 TOPOLOGY: linear
11 MOLECULE TYPE: protein
12
13 US-08-882-046-7

```

[illegible]

QY	950	FCNNNGSCAAVAMC	SQKGRQVS	-----	CSCKKGYKGGDEYSC	-----	LEIDPCA	----	993			
Db	778	ICAAQTNDPCSP	HPCYNSGTCVDDGMWY	-----	RCECAPAGAP	----	DRININECQSSPCA	FGAT	834			
QY	994	-DGVNG	-GC	-----	HEHATCR	-----	MTGPGKHRCCK	SHYVDDV	-----	DCEPEOLDPR	----	1038
Db	835	CVDELNGTRCVC	PPGHSGAKCQEFSG	-----	RPLCTMGSSVIR	PDGAKMKDDC	NTCQCLNGRIA	891				
QY	1039	-----	CLQDNG	-----	QCHPRASCA	-----	DLYPDQTTVGVFHLRS	PLGQTKLTFDRAK	108			
Db	892	CSKYWCGPRL	LHKHSGSECSG	QSCIP	LDIDDC	CFWHPCTGVEGRSS	-----	SLOPVK	945			
QY	1083	EACAKEA	-----	ATIA	-----	TYN	-----	QLSTYQAKKHYLCS	1110			
Db	946	TKCTSDSY	YODNCANIT	TFTEKEMSPGL	TTENHCS	981						

```

RESULT 14
; Sequence 6, Application US/09068740A-6
; Patent No. 6337387
; GENERAL INFORMATION:
; APPLICANT: SAKANO, SEIJI
; APPLICANT: ITOH, AKIRA
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8447
; CURRENT APPLICATION NUMBER: US/09/068,740A
; CURRENT FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: JP 7-299611
; PRIOR FILING DATE: 1995-11-17
; PRIOR APPLICATION NUMBER: JP 7-311811
; PRIOR FILING DATE: 1995-11-30
; PRIOR APPLICATION NUMBER: PCT/JP96/03356
; PRIOR FILING DATE: 1996-11-15
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 6
; LENGTH: 1036
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-068-740A-6

```

	Query Match	5.1%	Score 399.5;	DB 4;	Length 1036;	
	Best Local Similarity	21.7%	Pred. No. 3.le-24;			
	Matches 229;	Conservative 74;	Mismatches 318;	Indels 435;	Gaps 68;	
QY	189	PCGVRTIITRACWLSLAHNAKPARGEVVMCALGTASVWDGYNNGTGTCOGCGLGFGTACE	248			
	:   :	:	:	:  :  :  :	:	
Db	196	PCBNRAICROG-----SPKHGSCL-----PDCRCOYGWOGLYCD	232			
QY	249	PTCEBKRYGIHCDOACSVCYHGRCSCQPLDGGSCDCDVGMGRVKCDEI-----TTDMC--NG	302			
	:	:	:	:  :  :  :	:	
Db	233	KCIPIHP-----GCYHIGICNE-----PMQCILETNMGSQLCKDLNKCYNHQPCLNG	279			
QY	303	TCHTSA-----NCLLDP--DGKASK-----CAAFPRNG	330			
	:  :	:	:	:  :  :	:	
Db	280	TCSNTGPPDKYQCCSCEBGISGPCNELAEHNACLSDPCHNRKSCKETSLGFECESPGWTG-P	338			
QY	331	TWCATAINAGETSNNGCSTFKADCKRTTPGNRVYCVSKAGTYG-----DGIYLEINPLENH	385			
	:  :	:	:	:  :  :	:	
Db	339	TESTNIIDDCSPNN---CSHGCTQDLYNGFR-CVCPPOMTGKTQLDLANEC-EAKFCV---	391			
QY	386	GCCDRNAECTQTGPNOAVONCILPKYTGDKYSLI-INVCILTNNGCCSPFAFCNYTEODOR	444			
	:  :	:	:	:  :  :	:	
Db	392	-----NASCKMLIASYCYDCLPGMMGN--CDIINDCL-----GCCQNDASRDLYNGIR	441			
QY	445	ICTCKPDYTGDIIVRGSIYIGELPNPSTSOYFOLQEHAAVELAGPGFTVFAPLSSSF	504			
	:  :	:	:	:  :  :	:	
Db	442	CICRPYGAUGH--GERDI-DECAASNPCLN-----GGHCQWEI---NRFLQCLCTPGSG	488			
QY	505	N-----HEP-----RIKW-----DOQGIMSVLRHYHVGCQOLLNLNK	539			

Db 489 NLCQLDIDYCEBNPCONGAOCYNRASDYFCCKREDEYEGKNCSHLDKDH---CR----- 537  
QY 540 VTTSATTLTLOGEVSISVSODTVEFINNEAKVLSSDIISTNGVTHVIDKLSPNLLITPKD 599  
Db 538 -TTPEEVIDSCYVAAAM-SNDT-----PEGVATISSNVCGR----- 570  
QY 600 ALGRVQLNLTVAANHGYTKFKSLIQDSGLSVITDSIHTPTVTPMTDKALEALPREQO 659  
Db 571 -----HG----- 572  
QY 660 DPLFNQDNKDKLSYKLFHVIRDSKALASDLPRASAMKTLQSGSELSVRCGTSDIGELFL 719  
Db 573 -----KCKSQSGKFTCCDNKG-----FT 591  
QY 720 NEQMRFIHGRGLFDVAVAYGIDCLLMNPTLGRCDTFTTFDIPBGCSGCIPTPKPLKS 779  
Db 592 G-----TYCHENI-----NDC-ESNPCRNG-----GTCT----- 614  
QY 780 KPRGVKK-KCIYNPLFRNRNVEGCONLCTVVIQTPTPCGHYMPDCA-----CPGGPDP 834  
Db 615 --DGVNSYKCI-----CSDGMEGAYCETININDC---SQNP 644  
QY 835 CNNRGMCRDLTPRMGQCLCHTGFNGTACELCMHGRFPDPCQPSRCSGSEHQC-DEG1TGS 893  
Db 645 CHNGGTCRDLYVNDP-YCDCKNMGKKTG---HSR-DSQCDEATCNGGTCTYDEG--DAF 696  
QY 894 ECLCETGWTAAASCDPTTAVFAVCTP-ACSVNATCTENN---TCVCNLNTEGSGITCTVVD 949  
Db 697 KCMCGMGTGTCN--IANNSSCLPRCHNGGTGVVNGSFTCYCKEGMEG-----P 746  
QY 950 FCKONNGCAKAAKCSQKTQVS-----CCKKGYKGBGYSC-----LEIDPCA----- 993  
Db 747 ICAQNTNDCSP--HPCVNSGTCDVGDMMWRCECAPGAGP--DCRININEGQSPCAFGAT 803  
QY 994 --DGVNG-GC-----HEHATCR-MTGRGKHKCECKSHYUGDV-----DEPERDLPRDR-- 1038  
Db 804 CVDENIGYRCVPRPGHSGAKCEVSG--RPTTMGSVLPDGAKKMDDCNTCCCLNGRLA 860  
QY 1039 -----CLQDNG--OCHPDASCA-----DLYFQDTTVGVFNHLSPLGYKLFDPDKAK 1082  
Db 861 CSKVMCGRRPCLLHKHSGSPGSGSCIPLLDQCFVNRCTGVGECHSS-----SLQPVK 914  
QY 1083 EACAKEA-----ATTA-TYN-QLSTAQAKYHLCS 1110  
Db 915 TKCTSDSYQDNCANITFTFNKEMMSPLGTLTEHICS 950

RESULT 15  
US-09-068-740A-7  
; Sequence 7, Application US/09068740A  
; Patent No. 6337387  
; GENERAL INFORMATION:  
; APPLICANT: SAKANO, SEIJI  
; APPLICANT: ITOH, AKIRA  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8447  
; CURRENT APPLICATION NUMBER: US/09/068,740A  
; CURRENT FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: JP 7-299611  
; PRIOR FILING DATE: 1995-11-17  
; PRIOR APPLICATION NUMBER: JP 7-311811  
; PRIOR FILING DATE: 1995-11-30  
; PRIOR APPLICATION NUMBER: PCT/JP96/03356  
; PRIOR FILING DATE: 1996-11-15  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 1187  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-068-740A-7

Query Match 5.1%; Score 399.5; DB 4; Length 1187;  
Best Local Similarity 21.7%; Pred. No. 3.8e-24;  
Matches 229; Conservative 74; Mismatches 318; Indels 435; Gaps 68;  
QY 189 POCVTRITTRACWLASLAINAKPARGEVKMCALGTASVWDVNGTGTCCGCGGFGTACE 248  
Db 196 PECNRAICRQGC-----SPKHGSKL-----PDCCKQYGMQGLYCD 232  
QY 249 TCEGKYGIHCDQACSCVHGRCSQGPLGDSQCDVWGRVACDMEI-----TTDNC--NG 302  
Db 233 KCIPLP-----GCVAGICNE-----PMQCLCETNMGQGLCDMDLNCGRHQCCLNG 279  
QY 303 TCHTSA-----NCLLDP-DGKASCK-----CAAGFRNG 330  
Db 280 TCSNTPRDKYCCSCEGYSGPCEIAEHNACLSDPCHNRSGKETSIAGECECSPMTG-P 338  
QY 331 TWCATAMCEISNGGCSPTADCKRTTPGRVRCVACAGYTG-----DG1CILEINPLENH 385  
Db 339 TCSNTINDCSPNN--CSHGTCQDLYNGRK-CVCPROMTKYTCQLDANEC-BAKPCV--- 391  
QY 386 GGCDRNAECTGTGPNQAVNCCLPKYTGDKVCSL-INWCLTNNGGCSPPAFCNYTEODOR 444  
Db 392 -----NAKCKNLLASLYCDDLPGMMGN--CDININDCL---GGOQNDASGRDLVNGVR 441  
QY 445 ICTCKRPDYTGDIYCRGSIYGBLPKNPSTSYFLOLQEHAYBELAGRPFTYFAPLSSSF 504  
Db 442 -CICPPGYAGDI--CERDI-DECAENPLN-----GHCONEI---NRFQCLCPTGFSG 488  
QY 505 N-----HEP-----RIKDW-----DQGLMSQVLYRHVYGGQQLLDNLK 539  
Db 489 NLCQLDIDYCEBNPCONGAOCYNRASDYFCCKREDEYEGKNCSHLDKDH---CR----- 537  
QY 540 VTTSATTLTLOGEVSISVSODTVEFINNEAKVLSSDIISTNGVTHVIDKLSPNLLITPKD 599  
Db 538 -TTPEEVIDSCYVAAAM-SNDT-----PEGVATISSNVCGR----- 570  
QY 600 ALGRVQLNLTVAANHGYTKFKSLIQDSGLSVITDSIHTPTVTPMTDKALEALPREQO 659  
Db 571 -----HG----- 572  
QY 660 DPLFNQDNKDKLSYKLFHVIRDSKALASDLPRASAMKTLQSGSELSVRCGTSDIGELFL 719  
Db 573 -----KCKSQSGKFTCCDNKG-----FT 591  
QY 720 NEQMRFIHGRGLFDVAVAYGIDCLLMNPTLGRCDTFTTFDIPBGCSGCIPTPKPLKS 779  
Db 592 G-----TYCHENI-----NDC-ESNPCRNG-----GTCT----- 614  
QY 780 KPRGVKK-KCIYNPLFRNRNVEGCONLCTVVIQTPTPCGHYMPDCA-----CPGGPDP 834  
Db 615 --DGVNSYKCI-----CSDGMEGAYCETININDC---SQNP 644  
QY 835 CNNRGMCRDLTPRMGQCLCHTGFNGTACELCMHGRFPDPCQPSRCSGSEHQC-DEG1TGS 893  
Db 645 CHNGGTCRDLYVNDP-YCDCKNMGKKTG---HSR-DSQCDEATCNGGTCTYDEG--DAF 696  
QY 894 ECLCETGWTAAASCDPTTAVFAVCTP-ACSVNATCTENN---TCVCNLNTEGSGITCTVVD 949  
Db 697 KCMCGMGTGTCN--IANNSSCLPRCHNGGTGVVNGSFTCYCKEGMEG-----P 746  
QY 950 FCKONNGCAKAAKCSQKTQVS-----CCKKGYKGBGYSC-----LEIDPCA----- 993  
Db 747 ICAQNTNDCSP--HPCVNSGTCDVGDMMWRCECAPGAGP--DCRININEGQSPCAFGAT 803  
QY 994 --DGVNG-GC-----HEHATCR-MTGRGKHKCECKSHYUGDV-----DEPERDLPRDR-- 1038  
Db 804 CVDENIGYRCVPRPGHSGAKCEVSG--RPTTMGSVLPDGAKKMDDCNTCCCLNGRLA 860  
QY 1039 -----CLQDNG--OCHPDASCA-----DLYFQDTTVGVFNHLSPLGYKLFDPDKAK 1082  
Db 861 CSKVMCGRRPCLLHKHSGSPGSGSCIPLLDQCFVNRCTGVGECHSS-----SLQPVK 914

OY 1083 EACAKEA-----ATA-TYN-OLSYAQKAKYHICS 1110  
| : : | : : | : : | : : | : : | : :  
Db 915 TKCTSDSYQDNCANITFTFNKEMSPGLTTEHICS 950

Search completed: June 17, 2002, 12:29:43  
Job time: 229 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 17, 2002, 12:28:29 ; Search time 133.39 seconds  
(without alignments)  
3776.027 Million cell updates/sec

Title: US-09-842-930A-2  
Perfect score: 7861  
Sequence: 1 SLPSLLTRLEQMPDYSIFRG.....WGHCGRPMRSGQATTVTPR 1431

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/2/paa/PCRTUS\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep:\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep:\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep:\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep:\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep:\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep:\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep:\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep:\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep:\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep:\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pep:\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep:\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pep:\*  
24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep:\*  
25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep:\*  
26: /cgn2\_6/ptodata/2/paa/US060\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7861	100.0	1431	1 PCT-US01-13403-2	Sequence 2, Appl1
2	7861	100.0	1431	22 US-09-842-930A-2	Sequence 2, Appl1
3	5969	75.9	1394	1 PCT-US01-13403-25	Sequence 25, Appl1
4	5175.5	65.8	1192	1 PCT-US01-13403-62	Sequence 62, Appl1
5	5110	65.0	1394	22 US-09-842-930A-25	Sequence 25, Appl1
6	4906.5	62.4	1895	26 US-60-230-445-1388	Sequence 1388, Ap
7	4447.5	56.6	1069	1 PCT-US01-13403-63	Sequence 63, Appl1

8	3762.5	47.9	897	1 PCT-US01-13403-64	Sequence 64, Appl1
9	2813	35.8	2157	1 PCT-US99-30462-2	Sequence 2, Appl1
10	2813	35.8	2157	18 US-09-466-778-2	Sequence 2, Appl1
11	2810	35.7	1460	1 PCT-US01-13403-60	Sequence 60, Appl1
12	2797	35.6	1458	1 PCT-US01-13403-61	Sequence 61, Appl1
13	2710.5	34.5	669	21 US-09-715-417A-16	Sequence 16, Appl1
14	2701.5	34.4	669	21 US-09-715-417A-12	Sequence 12, Appl1
15	2131	27.1	1082	1 PCT-US01-16450-2725	Sequence 2725, Ap
16	2131	27.1	1082	1 PCT-US01-16450A-2725	Sequence 2725, Ap
17	1989.5	25.3	518	26 US-60-230-445-1786	Sequence 1786, Ap
18	1897.5	24.1	1039	26 US-60-207-214-329	Sequence 329, App
19	1881	23.9	1065	26 US-60-213-846-977	Sequence 977, App
20	1502.5	19.1	353	1 PCT-US99-30462-11	Sequence 11, Appl1
21	1502.5	19.1	353	18 US-09-466-778-11	Sequence 11, Appl1
22	1487	18.9	315	21 US-09-715-417A-4	Sequence 4, Appl1
23	1318.5	16.8	545	21 US-09-715-417A-18	Sequence 18, Appl1
24	1264.5	16.1	381	21 US-09-715-417A-14	Sequence 14, Appl1
25	1254.5	16.0	1535	26 US-60-230-445-1666	Sequence 1666, Ap
26	1209	15.4	244	21 US-09-715-417A-6	Sequence 6, Appl1
27	1186	15.1	1431	26 US-60-207-583-641	Sequence 641, App
28	1067	13.6	334	21 US-09-715-417A-20	Sequence 20, Appl1
29	1024	13.0	687	26 US-60-228-515-1080	Sequence 1080, Ap
30	774	9.8	457	1 PCT-US99-30462-5	Sequence 5, Appl1
31	774	9.8	457	18 US-09-466-778-5	Sequence 5, Appl1
32	762	9.7	457	16 US-09-244-112-108	Sequence 108, App
33	762	9.7	457	21 US-09-774-639-108	Sequence 108, App
34	762	9.7	457	23 US-09-969-730-110	Sequence 110, App
35	731.5	9.3	676	26 US-60-230-445-1625	Sequence 1625, Ap
36	685	8.7	835	26 US-60-230-445-1787	Sequence 1787, Ap
37	497.5	6.3	2871	19 US-09-538-092-1076	Sequence 1076, Ap
38	496.5	6.3	2911	22 US-09-825-751A-68	Sequence 68, Appl1
39	496.5	6.3	2912	1 PCT-US01-08631-36761	Sequence 36761, A
40	486	6.2	2523	1 PCT-US99-15817-3	Sequence 3, Appl1
41	486	6.2	2523	12 US-08-899-222-3	Sequence 3, Appl1
42	486	6.2	2523	15 US-09-121-457-3	Sequence 3, Appl1
43	484.5	6.2	2090	26 US-60-213-177-839	Sequence 839, App
44	484	6.2	2141	10 US-08-676-993-4	Sequence 4, Appl1
45	484	6.2	2141	10 US-08-676-993A-4	Sequence 4, Appl1

## ALIGNMENTS

RESULT 1  
PCT-US01-13403-2  
GENERAL INFORMATION:  
APPLICANT: WEIGEL, PAUL H.  
APPLICANT: ZHOU, BIN  
APPLICANT: WEIGEL, JANET A.  
TITLE OF INVENTION: IDENTIFICATION AND USES OF A HYALURONAN RECEPTOR  
FILE REFERENCE: 618754/JEP/199,538  
CURRENT APPLICATION NUMBER: PCT/US01/13403  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/245,320  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: 60/199,538  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1431  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
PCT-US01-13403-2

Query Match 100.0%; Score 7861; DB 1; Length 1431;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SLPSLLTRLEQMPDYSIFRGYIIHYMLASAIISADATYVFPNNEAIENYTRKKATSLK 60  
|||||



Db	361	VCVCKAGTGGAGI	VCLEINPCELEHNGGCDR	NACSTQTPR	QWAVCNCLP	RYTG	GKVC	CSLI	420																																														
Qy	421	NVCLTNNGGCS	SPFAFCMYTQDOR	ICICR	PDYTGDI	VC	KGS	YIGELPK	NPSTSYF	FOL	480																																												
Db	421	NVCLTNNGGCS	SPFAFCMYTQDOR	ICICR	PDYTGDI	VC	KGS	YIGELPK	NPSTSYF	FOL	480																																												
Qy	481	QEHAVREL	AGPPTVAP	LPSSSE	NHPPRK	IKMD	DOOGL	MSQV	LR	YHVAC	COOLLD	LN	LV	540																																									
Db	481	QEHAVREL	AGPPTVAP	LPSSSE	NHPPRK	IKMD	DOOGL	MSQV	LR	YHVAC	COOLLD	LN	LV	540																																									
Qy	541	TTSAT	TLOGE	VSIS	VSQD	TVF	INNE	KVLS	DSI	IST	NGV	I	HV	IDK	LP	KNLL	TP	PKDA	600																																				
Db	541	TTSAT	TLOGE	VSIS	VSQD	TVF	INNE	KVLS	DSI	IST	NGV	I	HV	IDK	LP	KNLL	TP	PKDA	600																																				
Qy	601	LGAR	LOWL	TTVA	AAH	GGT	KTSK	LI	OSGL	LS	VTDS	I	HP	RV	YF	MP	TD	KA	LE	LP	RE	QD	660																																
Db	601	LGAR	LOWL	TTVA	AAH	GGT	KTSK	LI	OSGL	LS	VTDS	I	HP	RV	YF	MP	TD	KA	LE	LP	RE	QD	660																																
Qy	661	FLF	NOD	KDK	LSY	K	HN	V	RD	S	KAL	ASD	L	PRS	AS	M	K	T	LOG	S	E	L	S	R	C	T	G	S	D	I	G	E	L	F	L	N	720																		
Db	661	FLF	NOD	KDK	LSY	K	HN	V	RD	S	KAL	ASD	L	PRS	AS	M	K	T	LOG	S	E	L	S	R	C	T	G	S	D	I	G	E	L	F	L	N	720																		
Qy	721	EOM	C	R	F	I	H	R	G	L	E	D	V	A	Y	A	G	I	D	C	L	M	N	P	L	G	R	C	D	F	T	T	E	D	I	P	E	G	C	S	C	I	F	T	P	K	C	L	K	780					
Db	721	EOM	C	R	F	I	H	R	G	L	E	D	V	A	Y	A	G	I	D	C	L	M	N	P	L	G	R	C	D	F	T	T	E	D	I	P	E	G	C	S	C	I	F	T	P	K	C	L	K	780					
Qy	781	PKG	K	K	K	C	I	N	P	L	P	F	R	N	V	E	G	G	O	N	L	C	I	V	I	O	P	R	C	H	G	F	M	P	R	C	O	A	C	P	G	P	D	P	C	N	R	M	840						
Db	781	PKG	K	K	K	C	I	N	P	L	P	F	R	N	V	E	G	G	O	N	L	C	I	V	I	O	P	R	C	H	G	F	M	P	R	C	O	A	C	P	G	P	D	P	C	N	R	M	840						
Qy	841	C	R	D	L	Y	T	P	M	G	O	C	L	T	H	G	N	G	T	A	C	E	L	C	M	H	R	G	P	C	O	P	R	S	C	S	E	M	H	O	C	O	E	G	I	T	G	S	E	C	L	E	N	G	900
Db	841	C	R	D	L	Y	T	P	M	G	O	C	L	T	H	G	N	G	T	A	C	E	L	C	M	H	R	G	P	C	O	P	R	S	C	S	E	M	H	O	C	O	E	G	I	T	G	S	E	C	L	E	N	G	900
Qy	901	WT	AS	C	O	T	P	A	V	A	V	C	T	P	A	S	A	N	A	T	C	E	N	T	C	N	L	N	I	N	E	D	I	T	C	T	V	N	P	C	O	N	N	G	G	K	960								
Db	901	WT	AS	C	O	T	P	A	V	A	V	C	T	P	A	S	A	N	A	T	C	E	N	T	C	N	L	N	I	N	E	D	I	T	C	T	V	N	P	C	O	N	N	G	G	K	960								
Qy	961	V	A	K	S	O	K	G	T	O	V	S	C	K	K	A	G	Y	G	D	G	S	C	E	I	D	P	C	A	D	V	A	G	N	G	C	H	E	N	A	T	C	R	M	G	P	R	K	N	K	E	C	K	1020	
Db	961	V	A	K	S	O	K	G	T	O	V	S	C																																										

Query Match	Best local Similarity	Matches 1063	Conservative 120	Mismatches 190	Indels 12	Gaps 5
Query 23	75.9%	Score 5969	DB 1	Length 1394		
Db 1	IOYNLAAIEADATVYVAPNNAL	ENYIREKKVLSLEDYVLRHYVLEEKLLKNDLHNG	60			
QY 83	MHRETMLEFSYLLAFELNDOLYVNEAP	INTYNNVATDKGVHIGLEKRVLETOKNRCDNNDT	142			
Db 61	MHRETMLEFSYFLSFLHNDOLYVNEAP	INTYNNVATDKGVHIGLCKVLEIQNRCDNNDT	120			
QY 143	IIVRECGKCSQAPCLETKPL	RETREKCIYSIFMGKRSVFTGCGPOCVATTITRACW	201			
Db 121	TIIRRCRTCSSELCPEFTKSLGNEKRRCI	ITYSFMGRITLFIICQPCVTRVITRECC	180			
QY 202	LASLHANNKPAEGEKMALGTA	SWDGVNGTGCOCGSGFNGTACERTESKYGICHQD	261			
Db 181	AGFPGQCGPCGNANONN	CFGNGICLDGVNGVGECEGEGFSGTACETCEBKYGICHQD	240			
QY 262	ACSCVHGRCSGPGLDGS	CCDVGNRGVKCDMEITTDNCGTCHTSANCLDPDGKASCK	321			
Db 241	ACSCVHGRCNOGRLDGS	CCDVGMRGVCHDATTEDMNCGTCHTSANCLTNSDGTASCK	300			
QY 322	CAAGRGNGTCTAI	INACETSMGCGSTKADCKRTTPGNVVCYCKAGYGGDGIYCLINPC	381			
Db 301	CAAGGCGNGTCTAI	NACETISMGGCSAKADCKRTTPGRVCRCKAGYGGDGIYCLINPC	360			
QY 382	LENHGGCGRNAECQTGTG	NOAVCNCLPXYTGGKVCSLINCLTNNGGCSPPAFCNYTEQ	441			
Db 361	LENHGGCGKMAECQTGTG	NOAACNCLPAYTGGKVCCTLINCLTNNGGCSSEFALCNHTQ	420			
QY 442	DORICTCPDYTGDCIV	CRGSYIGGLPKNPSTSYQFFOLQELNAVBLAGPEFTYFAPLS	501			
Db 421	VERTCTCKPNYIGD	FTCRGSYIQLPKNPSTSYQFFOLQELHFNVDLVPGEFTYFAPLS	480			
QY 502	SSFNIEPRTKMDQGL	MSOYRYHYVAGCOOLNLNKLWTSATTLLOGEPVATISVOPV	561			
Db 481	AAFDKARVKMDKGL	MPYLRHYHYVACHOLLENKLKLSNATSLQGEPIYATISVSQSTV	540			
QY 562	FINNEAKVLSDDI	ISTNGVIVHDKLSPKNLLITPKDALGVLYONLTVAANHGYTKFS	621			
Db 541	YINNAKALISDDI	ISTNGIVHIDKLSPKNLLITPKDSGSLILONLTVAANNIGKFS	600			
QY 622	KLIDSGILSVYTD	SIHHPVYFEMPTDKALELPRQODFLPNOONKDKSLYLFHVYR	681			
Db 601	NLIQDSGLSVYTD	SIHHPVYLFEMPTDKALELPRQODFLPNOONKDKSLYLFHVYR	660			
QY 682	DSKALASDLP	PSASAKTTLQGSSELSVRCGTGSDIGELFTLEMQCGRFIIHRLLDVCVAYGI	741			

Db 661 DAKVAVDLPTSTAMKTLQSGSELVSKGAGRIDGLFLNGQTCRIYQRELLFDLGAVYGI 720  
QY 742 DCLLMNPITLGRCDPFTTFEDIPGEGSCITFPPKPKLSPKPGVKKKCYNLPFRANEG 801  
Db 721 DCLLIDPLLGRCDITFTTFEDASGEGSCVNTSPSPKSPKPGVKKCYNLPFRANEG 779  
QY 802 CONLTVVITQPRCHGFMPCQACPGGPDTPCNRNRCMDLTYPMQOCLHTGFGNGTA 861  
Db 780 CREGSLVITQIPROCKGYFGRDQACPGGPAPCNRNRCMDLTYPMQOCLHTGFGNGTA 839  
QY 862 CELCHHGFPGDCCPRSCSEHGQDEGTTGSEGLCEFTGWTAAASCDTPTAFVAVTPAC 921  
Db 840 CEMCPGRFPGDCLPCGSGDHGQCDGTTGSGQCLCEFTGWTGTPSCDTPAVLPAVCTPPCS 899  
QY 922 VHAATCTENNTGCVNLNTYGGDGTCTVVPCKONNGCAKVAKCSOGTGVSCSKKGYG 981  
Db 900 AHAATCKENNTGCVNLNTYGGDGTCTVVPCKODNGCAKVAKCSOGTGVSCSKKGYG 959  
QY 982 DGYSCIEIDPCADGVNGGCHGHAATCRMTGPKHKCECKSHYVGDGDEPEOLDPLDRCLQ 1041  
Db 960 DGHSTETIDPCADGLNGGCHGHAATCRMTGPKHKCECKSHYVGDGLNCEPEOLDPLDRCLQ 1019  
QY 1042 DNGOCHPAPASCADLYFODTTVGVFHLRSPLOYKLTLPDKAKAKAKAEATITVYNOLSYA 1101  
Db 1020 DNGOCHADAKCVDLHFODTTVGVFHLRSPLOYKLTLPDKAKAKAEATITVYNOLSYA 1079  
QY 1102 OKAKYHLCAGMLGSGRAVPTTYASOKGANYGVIVDYGSRANKSEMDVFCYMKDYN 1161  
Db 1080 OKAKYHLCAGMLGSGRAVPTTYASOKGANYGVIVDYGSRANKSEMDVFCYMKDYN 1139  
QY 1162 CTCRAGYVGDGFCSGNLLQVLMSPSLTNLTLEVLAFSKSSARQOAFKHLTDLSTIRCT 1221  
Db 1140 CTCRAGYVGDGFCSGNLLQVLMSPSLTNLTLEVLAFSKSSARQOAFKHLTDLSTIRCT 1199  
QY 1222 LEVPONSGLPKSKLSGNDIEHLLNVNVEYNDLVNGFTLNGSOLLTFEESODQLH- 1280  
Db 1200 LEVPONSGLPKSKLSGNDIEHLLNVNVEYNDLVNGFTLNGSOLLTFEESODQLH- 1259  
QY 1281 OETRVVDGKSLIOMDIIANGILHISEPLRAPPTAATPAHSGLTGTFCAVAVLTGATA 1340  
Db 1260 TETRVVDGKSLIOMDIIANGILHISEPLRAPPTAATPAHSGLTGTFCAVAVLTGATA 1319  
QY 1341 LAAYTFRLKORTTGFOPRDKRTMLSMILLAS-----SSPRTISOTLCMRPQRHQPSP 1394  
Db 1320 LAAYTFYFRNRTTGFOPRDKRTMLSMILLAS-----SSPRTISOTLCMRPQRHQPSP 1376  
QY 1395 VPPSQ 1399  
Db 1377 FTDSE 1381

RESULT 4  
PCT-US01-13403-62

Sequence 62, Application PC/TUS0113403  
GENERAL INFORMATION:  
APPLICANT: WEIGEL, PAUL H.  
APPLICANT: WEIGEL, JANET A.  
TITLE OF INVENTION: IDENTIFICATION AND USES OF A HYALURONAN RECEPTOR  
FILE REFERENCE: 618754/JP/199,538  
CURRENT APPLICATION NUMBER: PCT/US01/13403  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/245,320  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: 60/199,538  
PRIOR FILING DATE: 2000-04-25  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 62  
LENGTH: 1192  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US01-13403-62

Query Match 65.8%; Score 5175.5; DB 1; Length 1192;  
Best Local Similarity 77.7%; Pred. No. 0;  
Matches 914; Conservative 105; Mismatches 146; Indels 11; Gaps 4;  
QY 231 NGTTCGCGGLGFGNGTACETCTGKGTGHCDDQACSCVHGRCSGPLGDDSCDDCVGMRGVK 290  
Db 8 DGTGVCCEGEGFSGTACTCTEGKYGHCDDQACSCVHGRCMOFGJGDDSCDDCVGMRGVH 67  
QY 291 CDMETITNMGCTHTSANCILIDPDGKASCKCAAGFRNGTVCCTAINTCETNSGCTSKA 350  
Db 68 CDMATETNMGCTHTSANCILIDPDGKASCKCAAGFRNGTVCCTAINTCETNSGCTSKA 127  
QY 351 DCKRTTPGNRVCVCAGYTGDIYVLEINPCLENHGGCDRNAECTOTGPBNAVNCLEPRY 410  
Db 128 DCKRTTPGNRVCVCAGYTGDIYVLEINPCLENHGGCDRNAECTOTGPBNAVNCLEPRY 187  
QY 411 TGDGVVCSLINVCLTNGGCSPPAFNCNTTEDOQICTCTKPDYTGIGYICRGSITGELRKN 470  
Db 188 TGDGVVCSLINVCLTNGGCSPPAFNCNTTEDOQICTCTKPDYTGIGYICRGSITGELRKN 247  
QY 471 PSTSQYFQLOEHAVRELAPGPFVFAFLSSFNHPEPRKDPMDQGLMSQVLRVHYVC 530  
Db 248 PSTSQYFQLOEHAVRELAPGPFVFAFLSSFNHPEPRKDPMDQGLMSQVLRVHYVC 307  
QY 531 OQLLDNKLKVTTSATTLQGEPIVSIVSODTVEINNEKAVYSSDIITSTNGVHIVDKLSP 590  
Db 308 OQLLDNKLKVTTSATTLQGEPIVSIVSODTVEINNEKAVYSSDIITSTNGVHIVDKLSP 367  
QY 591 KMLLTPKDAAGRLVONLTVAANHGYKESKLIDSGSLVITPSIHPPVAVFVPTDKA 650  
Db 368 KMLLTPKDAAGRLVONLTVAANHGYKESKLIDSGSLVITPSIHPPVAVFVPTDKA 427  
QY 651 LEALPEEODFLFNODNKKDKLSYLFHYIRDSKALASDLPSRASMKTLOGSELSRYCT 710  
Db 428 LEALPEEODFLFNODNKKDKLSYLFHYIRDSKALASDLPSRASMKTLOGSELSRYCT 487  
QY 711 GSDIGELFLNEOMCRPIHRLGLFDVGVAYIDCLLMNPITLGRCDITFTTFIDIPGEGSCIT 770  
Db 488 GSDIGELFLNEOMCRPIHRLGLFDVGVAYIDCLLMNPITLGRCDITFTTFIDIPGEGSCIT 547  
QY 771 FTRPKCPLSKPKGVKKKCYNLPFRANVEGCONLCTVVIQPRCHGFMPCQACPGG 830  
Db 548 FTRPKCPLSKPKGVKKKCYNLPFRANVEGCONLCTVVIQPRCHGFMPCQACPGG 806  
QY 831 PDPCCNRRGMDLTYPMQOCLHTGFGNGTACELCHHGRFGDCCPRSCSEHGQDEGTT 890  
Db 607 PDPCCNRRGMDLTYPMQOCLHTGFGNGTACELCHHGRFGDCCPRSCSEHGQDEGTT 666  
QY 891 GSGECLCEFTGWTAAASCDTPTAFVAVTPACSVHATCTENNTGCVNLNTYGGDGTCTVVPF 950  
Db 667 GSGECLCEFTGWTAAASCDTPTAFVAVTPACSVHATCTENNTGCVNLNTYGGDGTCTVVPF 726  
QY 951 CKONNGCAKVAKCSOGTGVSCSKKGYKGDYSCIEIDPCADGVNGGCHGHAATCRMTG 1010  
Db 727 CKONNGCAKVAKCSOGTGVSCSKKGYKGDYSCIEIDPCADGVNGGCHGHAATCRMTG 786  
QY 1011 PKHKECKSHYVGDGDEPEOLDPLDRCLQDNGOCHHADAKCVDLHFODTTVGVFHLRSP 1070  
Db 787 PKHKECKSHYVGDGDEPEOLDPLDRCLQDNGOCHHADAKCVDLHFODTTVGVFHLRSP 846  
QY 1071 LQGYKLTLPDKAKAKAEATITVYNOLSYAKAKYHLCAGMLGSGRAVPTTYASOKG 1130  
Db 847 LQGYKLTLPDKAKAKAEATITVYNOLSYAKAKYHLCAGMLGSGRAVPTTYASOKG 906  
QY 1131 GANVVGIVDYGSRANKSEMDVFCYMKDYNCTCKAGYVGDGFCSGNLLQVLMSPSLT 1190  
Db 907 GANVVGIVDYGSRANKSEMDVFCYMKDYNCTCKAGYVGDGFCSGNLLQVLMSPSLT 966  
QY 1191 NFLTEVLAFSKSSARQOAFKHLTDLSTIRGTLFVPPONSGLPKSKLSGNDIEHLLNVN 1250  
Db 967 NFLTEVLAFSKSSARQOAFKHLTDLSTIRGTLFVPPONSGLPKSKLSGNDIEHLLNVN 1026

[illegible][illegible]

Db	475	VFAFLSAPFBEARVDMOKYGLMPOVLKTHVACHQJLLEMLKLLSNATSLQGEPIYIS	534
QY	556	VSDQFVFINNEAKVLSDDIISTNGVTHVIDKLSPKNLLITPKDALGRVLOLTVVAHN	615
Db	535	VOSSTAVINNKAKLISDDIISTNGVIHIDKLSPKNLLITPKDSNGRIOMLITLATIN	594
QY	616	GYIKFKSLIODSLLSVITDSIHFTPTVMPDKLAEALPREDOOFLENODMKJLSYL	675
Db	595	GYIKFSMLIODSLLSVITDPIHTPTVLEWPTDQALHAPADQOOFLENODMKJLEYL	654
QY	676	KFHNIIRSKALASDLPRASAMKTLQSELSVBCGNSDGEIFLNDEOMCFIHRILBDV	735
Db	655	KFHNIIRAKVLADLPJSTAMKTLQSELSVBCGNSGARGDGLFNLGQICRIYQRELLFDL	714
QY	736	GVAVGIDCLMLNPJLGRCDTFTTEDIPBGSCIFTEPKOPLSKSPKGVKKKCIYNLPF	795
Db	715	GVAVGIDCLLIDPTLQGRCDTFTTPRASEGSCSVWTPSCPRWSPKGVKOKCIYN-LPF	773
QY	796	RRNVEGONLCTVAVIOTPRCHGUEFMPDQACRGGDPPTCNNRGKCRDLYTPMGQCLHT	855
Db	774	KRLNEGCREKCSLVIOIPRCKCKYFERDQACRGDAPCNNRGGLDQYSAITGCKCNT	833
QY	856	GFNQTACELCHGRFPRDQPRSCSEHGGCQDEGIIQSGBJLCEGTMTAASCDTPPAEAV	915
Db	834	GFNQTACEMCPGRFGRDCLPRCCSDHGGCQDDGIIQSGQJLCEGTMTGSCDTQVLAIV	893
QY	916	CTPACSHATCTENNNTVCNLTAEGRGITCTYVVDPCKONGGCAVAAKCSQKQTOVSSC	975
Db	894	CTPSCSAHACCKENNTCECMLDYEGBGIICTYVVDPCKONGGCAVAVARCSQKQTVSSC	953
QY	976	KKGKAGGAGYGCIEIDPCADGVNCGCNEHATCKMTGFKHCKSCSHYVGDGVCDEPOLP	1035
Db	954	QKGKAGBGHSGTEIDPCADGSLNCGCNEHATCKMTGFKHCKSCSKHYVGDGLNCEBOLP	1013
QY	1036	LDRCLQDNGCGRPDASCADLYFODTTPVGVFHLRSLPGQKILTFDKAKERAKEATITATY	1095
Db	1014	IDRCLQDNGCCHADACVLDLHFQDDTPVGVFHLRSLPGQKILTFDKAREKANEATIMATY	1073
QY	1096	NOLSYAKAKYHLCSAGMLSGGVAAPTTYASQKGCANVGIYDVGSRANKSEMDVQCY	1155
Db	1074	NOLSYAKAKYHLCSAGMLLETGVAAPTPTAFASQKGCAGVGIYDVGPRPNKSEMDVQCY	1133
QY	1156	RMKDYNCTCKAGYVGDGFSQSGNLLQVLMSFSLTNFLLEVLAFKSSASGAQFLKHLTD	1215
Db	1134	RMKDYNCTCKVGVVGDGFSQSGNLLQVLMSFSLTNFLLEVLATYNSASRGAQFLLEHLTD	1193
QY	1216	LSIRGTLFVNQNSGLRPNKSLSGRDIENHLTVNVNVSFYNDLVNGTFLRLMGSQLITFS	1275
Db	1194	LSIRGTLFVNQNSGLRENETLSGRDIENHLTVNSMFEFYNDLVNGTTLQRLGSLKLLITAS	1253
QY	1276	ODOLH-OETREFVNGRSLTQMDIIAANGSLIHIISEPLRAPRPTAATAHNSLGTGFCVAVL	1334
Db	1254	QDPLQPTETREFVNGRATLQMDITFASNGSLIHIVLSRPLKAPAPAVUTLTHGGLAGIFFAILL	1313
QY	1335	VTGAIALAAVSFLRKQRTGTGFORQDKRPLKMSWLLAS-----SSPRIQTLICMPQRR	1368
Db	1314	VTGAVALAAVSYFRINRRTIGFHFESSEEDIIIVAAALGQOEPENISNPIYESTTSAPRE--	1371
QY	1389	HPOSPTVTPSQ 1399	
Db	1372	-PSYDPEPTDSE 1381	

```

RESULT      6      US-60-230-445-1388
: Sequence 1388, Application US/60230445
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
: TITLE OF INVENTION: AND USES THEREOF
: FILE REFERENCE: CLO00765

```



Db 66 LEINPCLENHGCGDKNAECTGPNQAAACNCLPAYTGDKVCTLLINVCITKNGCSEFAI 125  
QY 436 CNYTEODORICTCKPDYTGDIWCRGSYIGELPKNPSTSOXFEOLOEHAVERLACGPFT 495  
Db 126 CNHGTQVVRTCTCKPNYIGDFTGKSGIYQELPKPKPTSQFEOLOEHPVADLVPGPFT 185  
QY 496 VFAPLSSSFNHPRIKMDQGLMSQVLRHYHVGCQOOLLNKLKVTTSATTLQGEPAVIS 555  
Db 186 VFAPLSAFADEARVKMDKYGMLPQVLRHYHVACHQELLEMLKLSNATSLQGEPIYIS 245  
QY 556 VSQDTVFNNNAKYLSSDITSTNGVIHVIDKLSPKLLITPKALGVLQNLTTVAANH 615  
Db 246 VSQSTVYNNNAKIISSDIISTNGIVHVIDKLSPKLLITPKDMSGRILOMLTTLATNN 305  
QY 616 GYTFKSKLIQSGSLVTTSDIHPFVTFWPTDKALEALPPQOOFELFNODMKDLKSYL 675  
Db 306 GYIKFSNLIQSGSLVTTDPIHTPVTLEFWPTDQALHALPAEQDFFLNQDNKDLKEYL 365  
QY 676 KFHVIRDSKALASDLPRSASMKTLQSGSELSVRCGTGSDIGELFLNEQMCRIHRLFLDY 735  
Db 366 KFHVIRDAKVALVDLPSTAMKTLQSGSELSVRCGAGRDIQGLFLNGQTCRIYQRELLFDL 425  
QY 736 GVAAGIDCLLMPILGKCDFTFTFDIGECGSCITFTKPCPLKSPKVKKKCIYNLPDF 795  
Db 426 GVAAGIDCLLIDPILGKCDFTFTFDASGECGSCVNPFCSPKWSKPKGVKOKCLYN-LPF 484  
QY 796 RRNEGCONLCTVVIOTPRCHGYFMPPCOACPGGPDTPCNRNGCRDLTPRMOGCLHT 855  
Db 485 KRNLBEGCERSLVYQIPROCKGYFRDCCQACPGGPDAPCNRNGVCLDOYSATGCECKNT 544  
QY 856 GFNGTACELCMHGRFGPCOOPRSCSEHQCDEGITGSGECLETGWTAAACDTPPAVEAV 915  
Db 545 GFNGTACEMCPGRGPRCLPCGSGDHQCDDGITGSGQCLCTGWTGPRSDTOAVLPAV 604  
QY 916 CTPACSVATCTENNTCYCINLYEBGDITCTVVPFCQONNGCAKVAACSOKEOVSCSC 975  
Db 605 CTPPCSAHATCKENNTCCINLYEBGDITCTVVPFCQONNGCAKVAACSOKEOVSCSC 664  
QY 976 KKGKGGDYSCIEIDPCADGVNGGCHEHATCRMTGPKHKCECKSHYVGDDGDCPEOLD 1035  
Db 665 QKGTGKGDSHCTEIDPCADGLNGGCHEHATCKMTGPKHKCECKSHYVGDDGDCPEOLD 724  
QY 1036 LDRCLQDNGQCHPADSCADLYFODTTVGVFHLRSPLOQYKLTLPDKAKEACAEATTIATY 1095  
Db 725 IDRCLODNGQCHADAKCYDLHFODTTVGVFHLRSPLOQYKLTLPDKAREACANEATMTATY 784  
QY 1096 NQLSYAAQAKYHILCSAGMLESGRAVYPTTYASQKCGANVGIIVDYSRANKSEMDVFCY 1155  
Db 785 NQLSYAAQAKYHILCSAGMLETGRVAYPTAFASQNGSGVGIIVDYGPRPNKSEMDVFCY 844  
QY 1156 RMKVYNTCKRAGYVGDGFGSGCNLLQVLMSPSLTNFLTEVLAFSSKARQAFELKHLTD 1215  
Db 845 RMKG-----SAGLFQOLLSRPTCS-----KRPDD 868  
QY 1216 LSLRGTLFVPONSGLPNGKSLSGRDIENHLLNVNVSFYNDLVNGTFLRTMLGSQLITFS 1275  
Db 869 LSLRGTLFVPONSGLENETLSGRDIEHLLANVSMEFYNDLVNGTTLQTRLSKILLTAS 928  
QY 1276 ODOULH-QETRRVVDGRSTIIQOMILIAANGILHITSEPLRARPPLAATAHSGLTGTICAVYL 1334  
Db 929 QDPLQPTETRRVVDGRSTIIQOMILIFASNGILHIVISRLKAPARPVTLTHGIGAGJFAFILL 988  
QY 1335 VTGAIALAAYFYRLKORTTFORFODKRTLMSTMULAS-----SSPRISOTLCMRPDR 1388  
Db 989 VTGAVALAAYFYRLNRTTIGFQHFSESEEDINVALALGQOEINISNPLYESTTSAPPE-- 1046  
QY 1389 HPQSPVTPSQ 1399  
Db 1047 -PSYDPFDSE 1056

RESULT 8  
PCT-US01-13403-64

Sequence 64, Application PC/TU0S0113403  
GENERAL INFORMATION:  
APPLICANT: WEIGEL, PAUL H.  
APPLICANT: ZHOU, BIN  
APPLICANT: WEIGEL, JANET A.  
TITLE OF INVENTION: IDENTIFICATION AND USES OF A HYALURONAN RECEPTOR  
FILE REFERENCE: 618754/JP/199,538  
CURRENT APPLICATION NUMBER: PCT/US01/13403  
CURRENT FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/245,320  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: 60/199,538  
PRIOR FILING DATE: 2000-04-25  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 64  
LENGTH: 897  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US01-13403-64

Query Match 47.9%, Score 3762.5; DB 1; Length 897;  
Best Local Similarity 76.1%; Pred. No. 6,6e-270;  
Matches 676; Conservative 87; Mismatches 114; Indels 11; Gaps 4;

QY 519 MSQVLRHYHVGCQOOLLNKLKVTTSATTLQGEPAVISVSQDTVFNNNAKYLSSDIISTN 578  
Db 1 MPOVLRHYHVACHQELLEMLKLSNATSLQGEPIYISVSQSTVYNNNAKIISSDIISTN 60  
QY 579 GVHVIDKLSPKLLITPKDALGVLQNLTTVAANHGYTFKFSKLIQSGSLVTTSDIH 638  
Db 61 GVHVIDKLSPKLLITPKDMSGRILOMLTTLATANNNGYIKFSNLIQSGSLVTTDPIH 120  
QY 639 TPVTFWPTDKALEALPEQDFFLNQDNKDLKSYLKFHVIRDSKALASDLPRSASMKT 698  
Db 121 TPVTLFWPTDQALHALPAEQDFFLNQDNKDLKEYLKFHVIRDAKVALVDLPSTAMKT 180  
QY 699 LQGSSELSVRCGTGSDIGELFLNEQMCRIHRLFLDYGVAAGIDCLLMPILGKCDFTFT 758  
Db 181 LQGSSELSVRCGAGRDIQGLFLNGQTCRIYQRELLFDLCVAAAGIDCLLIDPILGKCDFTFT 240  
QY 759 TFDIGECGSCITFTKPCPLKSPKGVKKKCIYNLPFRNNEGCONLCTVVIOTPRCHG 818  
Db 241 TFDASGECGSCVNPFCSPKWSKPKVKOKCLYN-LPFKRNLBEGCERSLVYQIPROCKG 299  
QY 819 YFMPDQACPGGPDTPCNRNGCRDLTPRMOGCLCTHGFNGTACELCMHGRFGPCOOPRS 878  
Db 300 YFGRCQACPGGPDAPCNRNGVCLDOYSATGCECKNTGFNGTACEMCPGRGPRCLPCG 359  
QY 879 CSEHQOCEBITGSGECLETGWTAAACDTPPAVAVCTPACSVATCTENNTCYCINLY 938  
Db 360 CSDHQOCEBITGSGQCLCTGWTGPRSDTOAVLPAVCTPPCSAHATCKENNTCCINLY 419  
QY 939 EGDGITCTVVPFCQONNGCAKVAACSOKEOVSCSKKGGDYSCIEIDPCADGVNG 998  
Db 420 ESDGITCTVVPFCQONNGCAKVAACSOKEOVSCSKKGGDYSCIEIDPCADGLNG 479  
QY 999 GCHENATCRMTGPKHKCECKSHYVGDDGDCPEOLDPLDRCLQDNGQCHPADSCADLYFQ 1058  
Db 480 GCHENATCRMTGPKHKCECKSHYVGDDGDCPEOLDPLDRCLQDNGQCHADAKCYDLHFQ 539  
QY 1059 DTTVGVFHLRSPLOQYKLTLPDKAKEACAEATTIATYNQLSYAAQAKYHILCSAGMLESGR 1118  
Db 540 DTTVGVFHLRSPLOQYKLTLPDKAREACANEATMTATYNQLSYAAQAKYHILCSAGMLETGR 599  
QY 1119 VAYPTTYASQKCGANVGIIVDYSRANKSEMDVFCYEMKDVNCTCKAGYVGDGFGSGCN 1178  
Db 600 VAYPTAFASQNGSGVGIIVDYGPRPNKSEMDVFCYEMKDVNCTCKAGYVGDGFGSGCN 659  
QY 1179 LLQVLMSPSLTNFLTEVLAFSSKARQAFELKHLTDLISRGTLFVPONSGLPNGKSLSG 1238  
Db 660 LLQVLMSPSLTNFLTEVLAVSNSSARQAFLEHLLTDLISRGTLFVPONSGLENETLSG 719

```

OY 1239 RDHHHTLNNVSFYNDLVNGFLRFLMTLSOLLTFESODOLH-OETRPVDSRSLIOMDII 1297
Db 720 RDIEHHLIANSMEFYNDLVNGVITLQTRVSGSKLLITASODPLOPETRPVDRGAILIOMDIF 779
OY 1298 AANGILHIISEPLRAPPTAATAHAHSGLTGTCICAVLVYGAIALAALAAYSYFRILKORTGFO 1357
Db 780 ASNGIHHVSRPLKAPPAVPYTLTHRTGLGANGIEFFAILLVGAVALAALAAYSYFRINRRTIGFO 839
OY 1358 RFDDKRLTMSWLAS-----SSFRISQFLCMRPQRHHRQSPVPPNQ 1399
Db 840 HFESEEDINVAALGKQOQENISNPLXYETTSAPPE--PSYDPFTDSE 884

RESULT 9
PCT-US99-30462-2
Sequence 2, Application PC/TUS9930462
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Novel Hyaluronan-Binding Proteins and Encoding Genes
FILE REFERENCE: PF487.PCT
CURRENT APPLICATION NUMBER: PCT/US99/30462
CURRENT FILING DATE: 1999-12-20
EARLIER APPLICATION NUMBER: 66/113,871
EARLIER FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2157
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US99-30462-2

```

Query Match	Similarity	35.8%	Score 2813;	DB 1.1%	length 2157;
Best Local	Similarity	40.1%	Pred. No. 7.8e+199;		
Matches	562;	Conservative	220;	Mismatches	554;
				Indels	66;
				Gaps	20;
QY	5	LTTRLEOMPDYSIFRKYIITHNYMLASIASADAYTFVFPNNEMAIENYIREKATSLKEDIL	64		
DB	718	LTQDLDIVAFSLFRLLQHHGLVPRLEAATATYIVFPNRSLE---AQGNSLHADPTV	774		
QY	65	RHYVVLGEKLLKNDLHNGHRETMIGFSLLAFPLRNDOLYNEAPINTNATDKGVH	124		
DB	775	RHHVVLGEALSMETLRKGGHRSNLGLPAWIMVFNHNSGQPEVNHVPLEGPMLEAPRSIL	834		
QY	125	GLEKVLKNCNDNDNLTIVREGCGSQOAPC---PLETKPLRETKCIYIYFMK	180		
DB	835	GLSGVLTVSSRCLSHMALEKREKVCNCTRRRCITQGPOLQPT---RKSCTYRSQFS	891		
QY	181	RSVFIGQPCQVRTIITRACWLSLAHNAKPAPEGYKMCALGTASYVMDGVNIGTCQGL	240		
DB	892	R---GCSYTAKKIKQVPPCCPGFFGLTECPBGGLGSGSGHQCDPFLGSGECHHE	947		
QY	241	GRNGTACETCTBKGKTIHCDQACSCYHNGCSQGLPBGSDCDDVCGKRGYKCMETITDMC	300		
DB	948	GRHGTAECVCELGRRGPNCTGVCDCAHGLCQBELQDGDGSCVGNVQGLRCQDKITISPPC	1007		
QY	301	NGTCHTSANCLLDPPDKSACKCAAFGRNGVYCTAINACETSGSGSTKADCKRTTPGR	360		
DB	1008	PKKCDPNACVODSAGASASTCACAGTSGMGICFSEVDPRAHGHGSGSPANCTKVAAPGR	1067		
QY	361	VCVCKAGYAGDGIYLEINPCLEHNHGGCORMAECTQTGFNOAVCNCLPKYTGDG-KVCSL	419		
DB	1068	TCTCGDGVGMDGLCOEINSLCHHGGCHHNAECIPTGGQVQVSCSREGSGGINTCEL	1127		
QY	420	INVCITLNNGCSGFACNCTEDQRICTCKPDT-GDGIVCRGSIGYGLPKNDSIQTQYF	478		
DB	1128	LDPGSKNNNGCSPYATCKSTGDDQPTCTQDTAHTVDDGLTCRARVGLLELLRKHAS--EF	1185		
QY	479	QIQEHAIVRELAGEPTVAP---LSSSNHNE--PIKIMDQOGLMSQYLRYHVACQOL	533		
DB	1186	SLRLLEYKELKGGPPTIVPNADLMSNLSDGLAIRAHQ-----LVFRYHVCRCRL	1240		

OY	534	LNDUKTTTSATLLOEAPISVISOQTPV INNEAKVLSDDIISTINGVHYIDKLLSPKNL	593
Db	1241	RSEDLLEOGYATATLSGHPRLPFSRESRESTYLNQFAVYSSDHEAVNGILHFDIVRLPREA	1300
OY	594	LITPKDALGVLYONLTTVAANHGYTKFSKLIODSCLLSVITDSITHTPYVWEPMDKALEA	633
Db	1301	LHMEPDDAPITPRRNRYAAAGFGYKLFSGLLLVADILRLPLEASHRPETMLMPDDAFA	1360
OY	654	LPEQODFLFNONKNIKASYLKFNHIDPSKALASDLPRSAWMKTLQSEELSTVPCGSGSD	713
Db	1361	LPPROQMWLTHEDHDKRLAAILRGHMITRVELVASDLNRLGLPKRMHETPLSEFSC-SRTR	1419
OY	714	IGELFLMEQCRPIHNGLELFDVGAVYIDCLMLNPTLGGRCDDTFTFDIP-GEQSCSIFT	772
Db	1420	PGELMWEDDARLYQHNHLPREGGLAVGIDQLLEPGLGARCIDHFETRLINTCSICGLE	1479
OY	773	PKCPKSKPKPGVKKK-----IYNLPR-----RNV-----EGQNCITV	808
Db	1480	PPCEGSOEGBSEACWREPKFEMTSPRLHSLGRLSRVWHPSLMRPGGLRGGRHRCVPT	1539
OY	809	VIOFPRCHCYFMPDOACAGGPDPTFCNNRGMCRODLYTPMGQCLCTHGFNTACELCMWG	868
Db	1540	TTMKPSCCPHYSSECQACPGGRSSPCSDRGYCMQMSGSGCCLCRSFATACELCAPG	1599
OY	869	RFGPDQPRSCSEHGQDEGISTGSECLCEGTMTAASCDTPPAVAVCTPACSVHATCTE	928
Db	1600	AFGPHQACRCHTVHGRCDDELGGSGCFDECMWTGPRCEVLELPVCTPCPAEAVACRA	1659
OY	929	NNTCVCMALNTEGOGITCTVYDFPCKOMNGGCANVAKCSQGTGYVSCSKKKYKGGYGSIE	988
Db	1660	GNSCESLGTEGGRKCTVADLQDDBHGCSSEHANCISOVTGWITCTCLPDTYEGWMSORA	1719
OY	989	IDPCADVNGCGCHEHATCRMTGPKHKECKSHYVODGVDC-EPEQLPLDRCLDNGSCH	1047
Db	1720	RNPFOTDGHRCGSHHANCJLSTGLNTRCECHAGYVODGJQCLSESEPPYVDCLQGRPCH	1779
OY	1048	PDASCADLYFDQTTGVGFHLRSPFGQYKLTLPDKACEAKKEATPATYTNOLSTYQAKAYH	1107
Db	1780	SDACTOLHFOEKRAGVFNHLOATSGPYGLNFSAEAMACEQACAVLASPQSLAQQOLEFH	1839
OY	1108	LCSAGMLESRVANPTTYASQOKGCANVVGIVDYGSRANKSEMDVDFCYRMDVNYCTKAG	1167
Db	1840	LCLMGWLANGSTAHPIVFPVADDCGNGRGVGLVSGARKNLSEKMDVYCFRRVDVACRANG	1899
OY	1168	YVGGDFS-CGQNLLOVLMSPFSLTNFLTVEAFYSKSSARQOAFKLHLDLSITRGLFVPO	1226
Db	1900	FVGGISTGCKLILDVLAATANSFTYCMJLGAYANNTQSGDLDFDLDELTYKTLFVPV	1959
OY	1227	NSGLPGKSLISGRDIEHHLTNVWVSFYNDLVNCTPLRTMGSQLLITTSQDYLHETFEV	1286
Db	1960	NEGVDWMTLSPMLDELHASNATLLLSAN-ASQCKLPAHSGSLIISDAGPDNSSWAPVA	2018
OY	1287	DG-----RSIILOMIDLIANGGLHIIISEPLRAP-----TATTAHSGLGIGTFCAVLVYTG	1337
Db	2019	PGTYVVSRIIYMDIMAFNGIITHALASLAPPOQAVLAXEAPVPAAGVAVLAAGALLG	2078
OY	1338	AIALAASYRFLKORTTGFQRF	1359
Db	2079	LVAQALYLRANGKPMGFGFSAF	2100
RESULT	10		
	US-09-466-778-2		
	Sequence 2, Application US/09466778		
	GENERAL INFORMATION:		
	APPLICANT: Haslings, Gregg et al.		
	TITLE OF INVENTION: Novel Hyaluronan-Binding Proteins and Encoding Genes		
	FILE REFERENCE: PF487		
	CURRENT APPLICATION NUMBER: US/09/466,778		
	CURRENT FILING DATE: 1999-12-20		
	EARLIER APPLICATION NUMBER: 60/113,871		
	EARLIER FILING DATE: 1998-12-23		



; NUMBER OF SEQ ID NOS: 37  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2157  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-466-778-2

Query Match 35.8%; Score 2813; DB 18; Length 2157;  
 Best Local Similarity 40.1%; Pred. No. 7.8e-199;  
 Matches 562; Conservative 220; Mismatches 554; Indels 66; Gaps 20;

```

QY 5 LTRLEQMPDYISIFRGYIIHYNLASISADAYTFVFNNEIENYIREKATSLKEDIL 64
DB 718 LLOQDLVPAISLRFELLQHNGLVQIEATYATTFVFNRSLE---AQGNSHLDADTV 774
QY 65 RYHVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLVYNEAPINTVNTATDKGYIH 124
DB 775 RHHVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLVYNEAPINTVNTATDKGYIH 834
QY 125 GLEKLELQKRNCDNNDITTYRGECKSCQAAPC---PLETKPLRETRKCIYSTYFMCK 180
DB 835 GLSGVLTVGSSRCLSHAEALREKCVNCTRRFCTQGFOLDTP---RKSCVYRSGSFS 891
QY 181 RSVFGCPQCVRTITRACWMLASLAHNAKPARGEVKMCALGTASVMDGVNGTCCOGL 240
DB 892 R-----GCYTTAKKIQVDDCCBGFPGTCTCEPCGGGLGVCVSHGCCODFLGSGRCHHE 947
QY 241 GPNGTACETEGKYGKIHCDQASCVHGRCSOGLGSDCDQDVGRVGRKCDMEITTDNC 300
DB 948 GFHGACVCECELGRYGPNCTGVCDAHGLQEGIGDGSVCVNVGMOGLRCDQKITSPQC 1007
QY 301 NGCTHTSANCLLDPGKASCCKCAAGFRNGYVCTAINACETISNGCSTKADCKRTTPGKR 360
DB 1008 PRKCDPNANCVDASGASTACAGYSGNGIFCSFVDPACAHGCGCSBPMANTKYAPQGR 1067
QY 361 VCVCAGATYGGIYCLLETINPCLEHNGCDRNAECTOTGPNOAVCNLKYKTDG-KVCSL 419
DB 1068 TCTGODGYMGDELCOELINSLIHNGGCHIHAECTPTGPPOVSCSCREYSDDGIRTEEL 1127
QY 420 INVCITNNGGSPFAFCVYTEDORICTCKPDYT-GDGIVCSTYGLPKNPSTSOYFF 478
DB 1128 LDPSKNNGCCPYATCKSTGDGRTCTDPAHTVGDGLTCARVGLLELRDKHAS--FF 1185
QY 479 QLOEHAVALAGPGFTYFAP---LSSPFNHE--PRIDMDQOGLMSOYLRIHYVCCOL 533
DB 1186 SLRLLEYKELGDDGPFTTFVHADLMSLSODELARIRAHRO----LVFRYHVVGCRL 1240
QY 534 LIDNLKVTTSATTOGEFVSISSODTFVINNEAKVLSDDIISTGVIIHVIDKLSPNL 593
DB 1241 RSEDLLEGVATLSGHLRSEREGSITYLNDFAVYSSDEHAUVGILHFDIVLLPDEA 1300
QY 594 LITRKDALGRVLQNTTVAAHNGYTKFSKLIDSGSLSVITDSIHTPVTVMPIDKALEA 653
DB 1301 LHMEPPDAPIRRRNVTAAGGFGYKIFSGLLKVALGILLPLREASHRPMTMLPTDAFRA 1360
QY 654 LPREODPFLFNODKDKSLKLFHVIDSKALASDLPRASMKTLQSSSELSVRCGTSD 713
DB 1361 LPPRQAMLYHEDHDKDAALIRGMIRNVALASDLPLNLPLRTMHGTPIISFSC-SRTR 1419
QY 714 IGEFLFNOMGRFTHRGILLPVGVAAGIDCLIMNPTLGCRDPTTFPIIP-CEGSCSIFT 772
DB 1420 PGLMVGDDARIVORHLPFEGGLAYGIDQLLEPPGLGARDCHFTETRLRLNTGICGLE 1479
QY 773 PKCPKSKPKGVKKK-----IYNLPFR---RNV-----EGCONLCTV 808
DB 1480 PPECEGSEOGSPACMFRYKFTSPRLSHLGLRSVWVHSLMGRPOGLGGRHNCVT 1539
QY 809 VIOFPCCHGVEMPDQACPGGPDTPCNRNMGMDLYTPMQOCLCHTFNFGACLCMHG 868
DB 1540 TYMKPSCCPGHYGECCQACPGGPPSPCDKGVCMGMSGSOCLCRSGFAGTACELAPG 1599
  
```

```

QY 869 RFPDQPRSCSEHGOCDEGITSGCELCETGWTAAACDTPPAVAVCTPACSVHATCTE 928
DB 1600 AFGPHQACRCYHGHGCDVGLGSSGCPDEBGTGRCEVOLELDQVCTPPCAPAVCA 1659
QY 929 NNTVCNLMYEGDGIYTCVDFCCKONNGCAGARVAKVCSQKQGVSCCKKGYKGDYSCIE 988
DB 1660 GNSCECSLGEEDGRVCTVADLCQDGHGSCSEHANCOSQVGTWVTCCLPDYEGDMSGRA 1719
QY 989 IDPCADGVNGGCHENATCMTGPBKHKCECKSHYVDGYDC-EPEODPLDRLQDNGGCH 1047
DB 1720 RNPCTDGHNGCSEHANCSTGINTRRCECHAGYVDQJCLEESPVPDRCLGQPPCH 1779
QY 1048 PASACADLYEOPDTGVFPLRSPDLGOYKLTIPKAKCAKAEATATYNDLSAOKAKXH 1107
DB 1780 SDAMKTDLHFQEKRAVCFHLOATSGPYGLNFSAEAKCAQGVILASFPOLSHAODUGFH 1839
QY 1108 LCSAGWLESGRVAYPTTYASQKGANVVGIVYGRANKSEMDVPCYEMKDVNCTCKAG 1167
DB 1840 LCLMGWLANGSTAHNPVFPVADCGNCRGIVSLGARKNLSEHMDACFPVQVADACRNG 1899
QY 1168 YVGDEFS-CSGMLQVLMSPSLTNFLTEVLAFSSKSSARGQAFKLHLDLSTRGTLFVQ 1226
DB 1900 FVGDDISTCNGKLLDVLATANFSTFYGMLLGYANATQGLDFLDDELTYKTLFV 1959
QY 1227 NSGLPQKSLSGRDIENHLTNVNSFYNDLVNGTFLRTMLGSOLLITFSQDQLHOETREV 1286
DB 1960 NEGFDNMFTLSGPMLSHASNAFTLSAN-ASQGLLPAHSGSLTITSDGPNSSAPYA 2018
QY 1287 DG-----RSLLQWDIAANGILHIIEPLRAPP-----TAATAHSGLGIGFCVAVLYTG 1337
DB 2019 PGTVVVSRILTVMDIAPNGIHALASPLAPPQPAVILAXEAPVPAAGVAILAAGALLG 2078
QY 1338 ATALAAVSTFRLKQRTTGFQRF 1359
DB 2079 LVAGALYLRARGKPMGFGFSAP 2100
  
```

RESULT 11  
 PCT-US01-13403-60  
 ; Sequence 60, Application PC/TUS0113403  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEIGEL, PAUL H.  
 ; APPLICANT: ZHOU, BIN  
 ; APPLICANT: WEIGEL, JANET A.  
 ; TITLE OF INVENTION: IDENTIFICATION AND USES OF A HVALURONAN RECEPTOR  
 ; FILE REFERENCE: 618754/Jp/199,538  
 ; CURRENT FILING DATE: 2001-04-25  
 ; PRIOR FILING DATE: 2000-11-02  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: 60/7199,538  
 ; PRIOR FILING DATE: 2000-04-25  
 ; NUMBER OF SEQ ID NOS: 64  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 60  
 ; LENGTH: 1460  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 PCT-US01-13403-60

Query Match 35.7%; Score 2810; DB 1; Length 1460;  
 Best Local Similarity 40.0%; Pred. No. 7.5e-199;  
 Matches 561; Conservative 222; Mismatches 546; Indels 72; Gaps 22;

```

QY 5 LTRLEQMPDYISIFRGYIIHYNLASISADAYTFVFNNEIENYIREKATSLKEDIL 64
DB 21 LLOQDLVPAISLRFELLQHNGLVQIEATYATTFVFNRSLE---AQGNSHLDADTV 77
QY 65 RYHVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLVYNEAPINTVNTATDKGYIH 124
DB 78 RHHVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLVYNEAPINTVNTATDKGYIH 137
  
```





```
; APPLICANT: Vernet, Corine
; APPLICANT: Fernandes, Elma
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-606
; CURRENT APPLICATION NUMBER: US/09/715,417A
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 60/166,336
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: 60/167,785
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 60/187,844
; PRIOR FILING DATE: 2000-03-08
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 669
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-715-417A-12
```

Query Match 34.4% Score 2701.5; DB 21: Length 669;

Best Local Similarity 78.6%; Pred. No. 2.9e-191;

Matches 470; Conservative 60; Mismatches 67; Indels 1; Gaps 1;

```
QY 509 RIKMDQGLMSQVLRHVYVGCQQLLDNLKVTTSATLQGEPSISVSQDFEINNEAK 568
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 QVKMDKYGCLMPQVLRHVYVACHOLLENLKLNATSLQGEPIVISQSTVIYINNAK 127

QY 569 VLSDIISTNGVIVHIDKLSPKNLLTPKDALGRVIONLTVAANHGYTFESKLIODSG 628
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 IISDIISTNGIIVHIDKLSPKNLLTPKDNISRIIONLTVAANNGYTFESKLIODSG 187

QY 629 LLSVITSIHPPVVFVPTQKALALPEODDLFNQDNKMLKSYLKEHVIIRDSKALAS 688
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 LLSVITSIHPPVVFVPTQKALALPEODDLFNQDNKMLKSYLKEHVIIRDSKALAS 247

QY 689 DLPRASMKTLQSGSELSVRCGTSGDIGELFLNEQMFIRHGLLEDVGVAVGIDCLLMP 748
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 DLPRASMKTLQSGSELSVRCGTSGDIGELFLNEQMFIRHGLLEDVGVAVGIDCLLMP 307

QY 749 TLGRCDTFTFDLPEGCGSCTFPPKCPKLPKGVKKKCLYNPLPRRNVGCONCLTV 808
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 308 TLGRCDTFTFDLPEGCGSCTFPPKCPKLPKGVKKKCLYNPLPRRNVGCONCLTV 366

QY 809 VIOPRCHGTFMPDCOACPGPPTPCNNRGMCRDLYTPMGOCLOHNGFNGTACELCWG 868
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 VIOPRCHGTFMPDCOACPGPPTPCNNRGMCRDLYTPMGOCLOHNGFNGTACELCWG 426

QY 869 RFGPDQPRSCGSEHGQDEGITYSGECLCETGWTAAASCDPTTAVFAVCTPACSVHATCTE 928
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 RFGPDQPRSCGSEHGQDEGITYSGECLCETGWTAAASCDPTTAVFAVCTPACSVHATCTE 486

QY 929 NNTCVMLNTEGDDITCTVYDFCKQNNGGCAKVAKSGKGTQVSCSKKGKGBGYSIE 988
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 NNTCVMLNTEGDDITCTVYDFCKQNNGGCAKVAKSGKGTQVSCSKKGKGBGYSIE 546

QY 989 IDPCADVNGGCHHATCRMTGPGKHKECKSHVVGDDVDPEPOLPDRCLQDNGOCHP 1048
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 IDPCADVNGGCHHATCRMTGPGKHKECKSHVVGDDVDPEPOLPDRCLQDNGOCHP 606

QY 1049 DASCADYFDDTIVGVPHLSPLQYKLLTPDKAKACAKAATATATYNOLSYAAQAKY 1106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 DAKVDLHFDQTTGVGVPHLSPLQYKLLTPDKAKACAKAATATATYNOLSYAAQAKY 664
```

```
RESULT 15
PCT-US01-16450-2725
; Sequence 2725, Application PC/TUS0116450
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA131PCT
```

```
; CURRENT APPLICATION NUMBER: PCT/US01/16450
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2820
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2725
; LENGTH: 1082
; TYPE: PRP
; ORGANISM: Homo sapiens
PCT-US01-16450-2725
```

Query Match 27.1% Score 2131; DB 1: Length 1082;

Best Local Similarity 41.6%; Pred. No. 1.4e-148;

Matches 430; Conservative 155; Mismatches 396; Indels 52; Gaps 16;

```
QY 370 GDGIYCLEINPCLENHGGCDNRNACTGTGPNQAVNCNLPKTYTGG-KVCSLINVCNTNG 428
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 GDGIYCLEINPCLENHGGCDNRNACTGTGPNQAVNCNLPKTYTGG-KVCSLINVCNTNG 61

QY 429 GCSPPAFNCNTEVDORICTCKPDYT-GDGIYCRGSISYGLPKNPSTSGYFQLOEHAVRE 487
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 GCSPPAFNCNTEVDORICTCKPDYT-GDGIYCRGSISYGLPKNPSTSGYFQLOEHAVRE 119

QY 488 LAGPPTVFAP---LSSSFNHE--PRIKMDQGLMSQVLRHVYVGCQQLLDNLKVTY 542
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 LAGPPTVFAP---LSSSFNHE--PRIKMDQGLMSQVLRHVYVGCQQLLDNLKVTY 174

QY 543 SATTLQGEPSISVSQDFEINNEAKVLSDIISTNGVIVHIDKLSPKNLLTPKDALG 602
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 SATTLQGEPSISVSQDFEINNEAKVLSDIISTNGVIVHIDKLSPKNLLTPKDALG 234

QY 603 RVLNLTVAAANHGYTFESKLIODSGILSVTDSIHPPVVFVPTQKALALPEODDL 662
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 RVLNLTVAAANHGYTFESKLIODSGILSVTDSIHPPVVFVPTQKALALPEODDL 294

QY 663 FNQDNKMLKSYLKEHVIIRDSKALASDLPRASMKTLQSGSELSVRCGTSGDIGELFLNEQ 722
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 FNQDNKMLKSYLKEHVIIRDSKALASDLPRASMKTLQSGSELSVRCGTSGDIGELFLNEQ 353

QY 723 MCRFIHGLLEDVGVAVGIDCLLMPNPTLIGRCDFTTFDIP-GEQSGCTFPPKCPKLPK 781
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 MCRFIHGLLEDVGVAVGIDCLLMPNPTLIGRCDFTTFDIP-GEQSGCTFPPKCPKLPK 413

QY 782 KGVKKKC-----LYNPLPRF-----RNV-----EGCONLCYVAVIOTPRCH 817
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 414 KGVKKKC-----LYNPLPRF-----RNV-----EGCONLCYVAVIOTPRCH 473

QY 818 GFMPDCOACPGPPTPCNNRGMCRDLYTPMGOCLOHNGFNGTACELCWGFRGPDQPR 877
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 GFMPDCOACPGPPTPCNNRGMCRDLYTPMGOCLOHNGFNGTACELCWGFRGPDQPR 533

QY 878 SCSEHGQDEGITYSGECLCETGWTAAASCDPTTAVFAVCTPACSVHATCTNNTVCYNLN 937
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 SCSEHGQDEGITYSGECLCETGWTAAASCDPTTAVFAVCTPACSVHATCTNNTVCYNLN 593

QY 938 YEGGICTCTVYDFCKQNNGGCAKVAKSGKGTQVSCSKKGKGBGYSIEIDPCADGVN 997
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 594 YEGGICTCTVYDFCKQNNGGCAKVAKSGKGTQVSCSKKGKGBGYSIEIDPCADGVN 653

QY 998 GGCSEHNAHATCRMTGPGKHKECKSHVVGDDVDPEPOLPDRCLQDNGOCHPDAACADY 1056
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 GGCSEHNAHATCRMTGPGKHKECKSHVVGDDVDPEPOLPDRCLQDNGOCHPDAACADY 713

QY 1057 FQDTTGVPHLSPLQYKLLTPDKAKACAKAATATATYNOLSYAAQAKYHLSAGWLES 1116
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 FQDTTGVPHLSPLQYKLLTPDKAKACAKAATATATYNOLSYAAQAKYHLSAGWLES 773

QY 1117 GRVAVPTTYAOKGCANVYGVYDGSRANKSEMDVPCYRKADYVNCCKAKAYVDGDS-C 1175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 774 GRVAVPTTYAOKGCANVYGVYDGSRANKSEMDVPCYRKADYVNCCKAKAYVDGDS-C 833
```

```
QY 1176 SGNLLQVLMSEPSLTNFLTVEVLAFSSKSSARQOAFELKHLJTLSTIRGTLFVPQNSGLPGNKS 1235
      :| | | | : : : | : : : | | | | | | | | :
Db 834 NGKLDVLAATANSTFYGMILGYANATQRGDLDELDELTYKTLFVPVNEGSPVDMKT 893
      | | | | | | | | | | | | | | | | | | | |
QY 1236 LSGRDIHHLITNNVVSFYNDLVNGTFLRTMIGSOLLITFSODLHQETREVDG----RSI 1291
      | | | | | : : : | | | | | | | | | | | |
Db 894 LSGPDELHASNATLISAN-ASOGKLLPAHSGLSLIISDAGPDNSSWAPVAPGTVVVSRI 952
      | | | | | : : : | | | | | | | | | | | |
QY 1292 LQMDITIANGLTIHISEPLRAPP-----TATTAHSGIGTGIFCAVVLVTGAIALAASY 1346
      :| | | | | | | | : : : | | | | | | | | :
Db 953 IWDIMAFNGIITHALASFLAPQOPQAVLAPPAEPVAGVAGAVLAGALLGLVAGALTYLR 1012
      | | | | | | | | | | | | | | | | | | | |
QY 1347 FRLKQRTTGFORF 1359
      | | | | | | | | | | | | | | | | | | | |
Db 1013 ARGKPMGFGFSAF 1025
      | | | | | | | | | | | | | | | | | | | |
```

Search completed: June 17, 2002, 12:33:29  
Job time: 300 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:28:54 ; Search time 29.51 Seconds  
(without alignments)  
4088.566 Million cell updates/sec

Title: US-09-842-930A-2  
Perfect score: 7861  
Sequence: 1 SLPSLLTRLEQMPDYSIFRG.....WGHGPDMSQATTVTPR 1431

Scoring table:  
BLSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 256022 seqs, 84314167 residues

Total number of hits satisfying chosen parameters: 256022

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/US06\_NEW.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US07\_NEW.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US08\_NEW.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US09\_NEW.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US10\_NEW.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW.COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW.COMB.pep:\*

Pred. No. is the number of results predicted by the chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6072	77.2	1986	US-10-123-962-2	Sequence 2, Appli
2	6072	77.2	2017	US-10-123-962-4	Sequence 4, Appli
3	1094	13.9	510	US-10-104-047-2580	Sequence 2580, Ap
4	762.5	9.7	482	US-10-104-047-3794	Sequence 3794, Ap
5	480.5	6.1	1111	US-60-360-039-5280	Sequence 5280, Ap
6	399	5.1	3507	US-60-360-039-5784	Sequence 5784, Ap
7	375.5	4.8	1584	US-60-360-039-6549	Sequence 6549, Ap
8	373.5	4.8	1106	US-60-360-039-5709	Sequence 5709, Ap
9	371	4.7	3051	US-60-360-039-5866	Sequence 5866, Ap
10	359.5	4.6	1372	US-60-360-039-5971	Sequence 5971, Ap
11	338	4.3	3494	PCT-US02-09671-697	Sequence 697, App
12	338	4.3	3494	PCT-US02-09671-703	Sequence 703, App
13	338	4.3	3594	US-10-150-821-4	Sequence 4, Appli
14	338	4.3	3623	PCT-US02-09671-698	Sequence 698, App
15	338	4.3	3623	PCT-US02-09671-701	Sequence 701, App
16	338	4.3	3623	PCT-US02-09671-702	Sequence 702, App
17	335	4.3	3672	US-60-360-039-6146	Sequence 6146, App
18	334.5	4.3	3571	US-10-150-821-2	Sequence 2, Appli
19	333	4.2	1808	US-60-360-039-5986	Sequence 5986, App
20	331	4.2	3712	US-10-108-605-103	Sequence 103, App
21	326.5	4.2	1429	US-60-360-039-5819	Sequence 5819, App
22	326.5	4.2	3298	US-60-365-384-181	Sequence 181, App
23	326.5	4.2	3332	US-60-365-384-182	Sequence 182, App
24	325	4.1	2199	US-08-793-273C-2	Sequence 2, Appli
25	324.5	4.1	1810	US-08-793-273C-4	Sequence 4, Appli
26	324.5	4.1	2201	US-10-100-912-2	Sequence 2, Appli

27	322	4.1	1529	US-09-766-511B-29	Sequence 29, Appli
28	315.5	4.0	1247	PCT-US02-08253-216	Sequence 216, App
29	312.5	4.0	1512	US-09-970-944-32	Sequence 32, Appl
30	307	3.9	2823	US-60-360-039-5220	Sequence 5220, App
31	307	3.9	2823	US-60-360-039-5221	Sequence 5221, App
32	306	3.9	1529	US-09-312-283B-396	Sequence 396, App
33	299	3.8	988	US-60-369-876-2	Sequence 2, Appli
34	297	3.8	3871	US-60-360-039-6538	Sequence 6538, App
35	297	3.8	3871	US-60-360-039-6539	Sequence 6539, App
36	295.5	3.8	1534	US-09-970-944-30	Sequence 30, Appli
37	295	3.8	1295	US-60-360-039-5820	Sequence 5820, App
38	294	3.8	1295	US-60-360-039-5821	Sequence 5821, App
39	294	3.7	2715	US-10-042-865-52	Sequence 52, Appl
40	292	3.7	1531	US-09-970-944-31	Sequence 31, Appli
41	292	3.7	2824	US-60-360-039-5166	Sequence 5166, App
42	289.5	3.7	737	US-10-121-062-38	Sequence 38, Appli
43	288.5	3.7	1557	US-60-360-039-6816	Sequence 6816, App
44	286.5	3.6	838	US-60-360-039-6445	Sequence 6445, App
45	284.5	3.6	3501	US-10-123-155-37	Sequence 37, Appli

ALIGNMENTS

RESULT 1				
US-10-123-962-2				
Sequence 2, Application US/10123962				
GENERAL INFORMATION:				
APPLICANT: Fiddle, Carl Johan				
APPLICANT: Gerhardt, Brenda				
APPLICANT: Zambrowicz, Brian				
TITLE OF INVENTION: Novel Human EGF-Family Proteins and Polynucleotides Encoding t				
FILE REFERENCE: LEX-0338-USA				
CURRENT APPLICATION NUMBER: US/10/123,962				
PRIOR FILING DATE: 2002-04-16				
PRIOR APPLICATION NUMBER: US 60/286,141				
PRIOR FILING DATE: 2001-04-24				
NUMBER OF SEQ ID NOS: 5				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 2				
LENGTH: 1986				
TYPE: PRT				
ORGANISM: homo sapiens				
US-10-123-962-2				
Query Match 77.2%; Score 6072; DB 6; Length 1986;				
Best Local Similarity 77.0%; Pred. No. 0;				
Matches 1083; Conservative 121; Mismatches 191; Indels 12; Gaps 5;				
QY	1	SLPSLLTRLEQMPDYSIFRGYIIHYNLASAIESADAYTVFPNNEAIENYIREKKATSLK	60	
DB	571	SLPSLLTRLEQMPDYSIFRGYIIHYNLASAIESADAYTVFPNNEAIENYIREKKATSLK	630	
QY	61	EDILRHVVHLGKILKNDLHNGMHRETMLGFSYLAFLRNOLQVYNEAPIVNTVATDK	120	
DB	631	EDILRHVVHLGKILKNDLHNGMHRETMLGFSYLAFLRNOLQVYNEAPIVNTVATDK	690	
QY	121	GVYHGLEKYLEIOTKNCNDNDITIVRGEGCKSCQAPCPLETKPL-REIRKCIYSIFENG	179	
DB	691	GVYHGLEKYLEIOTKNCNDNDITIVRGEGCKSCQAPCPLETKPL-REIRKCIYSIFENG	750	
QY	180	KKSVEFGCPQPCVRIITTRACMLASIAHNAKPAPEGVKKALGTASVWDGVNGTGCQGC	239	
DB	751	KRTLEFGCPQPCVRIITTRACMLASIAHNAKPAPEGVKKALGTASVWDGVNGTGCQGC	810	
QY	240	LEFNGACETCEGKYGKIHCDQACSVHGRCSOGPLGDSQCDQGVKGMCDMETTNN	299	
DB	811	BEFSGACTCEGKYGKIHCDQACSVHGRCSOGPLGDSQCDQGVKGMCDMETTNN	870	
QY	300	CNGCTHSANCLLPDGKASCKACAGFRGNGTVCTAINACETSNNGCSTKADCKRTTNG	359	
DB	871	CNGCTHSANCLLPDGKASCKACAGFRGNGTVCTAINACETSNNGCSTKADCKRTTNG	930	

QY	360	RVCCKRGYGGDGYVCLIEIMPCLENNHGGCRNACBQOTGNOVAQNCCLPXYTDGKVCIS	411
Db	931	RVCCKRGYTDGDIYVCLIEIMPCLENNHGGCRNACBQOTGNOVAQNCCLPXYTDGKVCIL	990
QY	420	INVCILTNNGGCSPPAFNCYNTEDODRITCKPDTYDGIYCRGSIYGEIPKPNSTSYFQ	479
Db	991	INVCILTNNGGCSPPAFNCYNTEDODRITCKPDTYDGIYCRGSIYGEIPKPNSTSYFQ	105
QY	480	LOEFAVELAGPGEYFAPRLSSGFNHERIKDMDOOGILMSOYLARHYVGGCOOLLIDNLK	539
Db	1051	LOEFAVELAGPGEYFAPRLSSGFNHERIKDMDOOGILMSOYLARHYVGGCOOLLIDNLK	111
QY	540	VTTSATLLOGEPPVSIYSVODTVEFINNAKVLSSDIISTNGVHVIDKLSPKMLITPKD	599
Db	1111	LISNATSLQGEPIYIYSVSQTHVYINNKAKITSSDIISTNGVHVIDKLSPKMLITPKD	117
QY	600	ALGAVLONLTTVAANHGYTFKSKLIDSGLLSYTDSIHPIYVYEMPTDKALEALPREQ	659
Db	1171	NSGRIELONLTTLANNNYIKFSNLIDSGLLSYTDPHPIYVLFMTDIALHALPAEQ	123
QY	660	DFELNQNOKDKLSYLFKFNHVRDSKALASDLPRASAKWTQSGELSVRCOTGSDIGELFL	719
Db	1231	DFELNQNOKDKLKYLFKFNHVRDQKVLAVDLPSTAKTITQSGELSVKCGAGNIDIDLFL	129
QY	720	NEOMCRFIHRLGLFDVGAIVGIDCLMLNPVLGGHCDTFTTFDIPGEGCSITFPKCPKLS	779
Db	1291	NGQCRIRVORELEDFDLCAVAGIDCLLIDLPVLGGHCDTFTTFDIPGEGCSVNTPSGPRMS	135
QY	780	KPKGVKKCITNPLPFRFRNNGCONLTLYIOPRCCHGFMPCQACPCGPTPCNNRG	839
Db	1351	KPKGVKKCKLYN-LPFRKNEBGCERSLSYIOTPRCKKGTFGDCCQACPGPRAPPCNNRG	140
QY	840	MCRDLYTPMGOCLOHTGFNGTACELCMHGRFRPDCQPRSCSEHGQCDDEGITSGECLSET	899
Db	1410	VCLIOYSATGCECKNTGFNGTACEMCPGRFPGDCLPCGSDHQQCDGDTTSGGQCLSET	146
QY	900	GWTASODTPIAVAVACSTPACSVNATCTENNVCVNLNTEGGDGTTCYVDFQCONNGCA	959
Db	1470	GWTGPSODTQAVLAACVCTPPCSAHATCKENNTCCNLDYGGDGTTCYVDFQCDONGCA	152
QY	960	KVACCSOKGTQVSCSCSKKGYKGDYSCEIEDPCADGVNGGCSHEHATCRMTGPKKHCECK	1011
Db	1530	KVACCSOKGTQVSCSCQKGYKGDHSGHSTIEDPCADGLNGGCSHEHATCKMTGPKKHCECK	158
QY	1020	SHYVGDDVDCERBOLPDLRCLQDNGGCHPRDASCADLYFQDTYGVFNHLSRPLQKLTED	107
Db	1590	SHYVGDDGINCERBOLPDLRCLQDNGGCHADAKCYDLHFQDTYGVFNHLSRPLQKLTED	164
QY	1080	KAKESCAKEATTIMTYMOLSYAOKAKKHLCSAGVLESGRAVAPPTYASQKCGANVGYD	113
Db	1650	KAREPCANEATTIMTYMOLSYAOKAKKHLCSAGVLESGRAVAPPARPSQMGSGVGYD	170
QY	1140	YGSRAANSSEMDVCEYRMKIDVNTCKAGYVGDDGFSGNNLLOYLMSFSLTNEFLTEYLAF	119
Db	1710	YGPBRNKSEMDVCEYRMKDYNCTCKAGYVGDDGFSGNNLLOYLMSFSLTNEFLTEYLAV	176
QY	1200	SKSSARQGAFLKHLTDLSIRCTPLVPONSGLPQMKSLISGDEIEHHLTNVAVSYNDLYNG	125
Db	1770	SNSSARQGAFLKHLTDLSIRCTPLVPONSGLGENETLISGDEIEHHLTNVAVSMFYNDLYNG	182
QY	1260	TEFLRTMLGSQLITFSQDOLH-QETRYVDGKSLIOMDIIAANGILHITSEPLAPPAAT	131
Db	1850	TEFLQTRILGSKILLITASODLPOTETRYVDGAILIOMDIIIPASNGIINHISRLAKAPAPVY	188
QY	1319	AAHSGLGTGICAVVLYTGAILAALAASTYFRLKQRTTGFQRFQDOKRTILMSULLAS-----	137
Db	1890	LTHGIGLGIGIFAILIYTGAVALLAASTYFRINRRTIGFOHRESEEDINVAALGKQOEPENI	194
QY	1373	SSPRTSOTLCMRPQRHRHQPSPVPPV 1399	
Db	1950	SNPLTESTTSAPPE---PSIDPTFDSE 1973	

```

RESULT      2
US-10-123-962-4
: Sequence 4, Application US/10123962
: GENERAL INFORMATION:
: APPLICANT: Friddle, Carl Johan
: APPLICANT: Gethardt, Brenda
: APPLICANT: Zambrowicz, Brian
: TITLE OF INVENTION: Novel Human EGF-Family Proteins and Polynucleotides Encoding t
: FILE REFERENCE: LIX-0338-USA
: CURRENT APPLICATION NUMBER: US/10/123,962
: PRIOR FILING DATE: 2002-04-16
: PRIOR FILING DATE: 2001-04-24
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 2017
: TYPE: PRT
: ORGANISM: homo sapiens
: US-10-123-962-4

```

	Query Match	Similarity	77.2%	Score	60.72:	DB	6:	Length	2017:
	Best Local	Similarity	77.0%	Pred	No. 0:				
	Matches	1083:	Conservative	121:	Mismatches	191:	Indels	12:	Gaps
Qy	1	SLP...LRL...EOM...P...D...SIF...RGY...II...HYN...LAS...E...SAD...A...T...V...P...NNE...AI...ENY...IRE...K...RATS...L...K	60						
Db	602	SLP...LRL...EOM...P...D...SIF...RGY...II...HYN...LAS...E...SAD...A...T...V...P...NNE...AI...ENY...IRE...K...RATS...L...K	661						
Qy	61	EDIL...H...V...L...G...E...K...L...K...N...D...L...N...H...M...H...E...T...M...G...E...Y...L...A...F...L...R...N...O...L...V...N...E...P...I...N...T...N...A...T...D...K	120						
Db	662	EDVL...Y...H...V...L...E...K...L...K...N...D...L...N...H...M...H...E...T...M...G...E...Y...L...A...F...L...R...N...O...L...V...N...E...P...I...N...T...N...A...T...D...K	721						
Qy	121	GV...H...L...E...V...L...E...H...O...K...N...R...C...D...N...D...F...I...Y...R...E...G...E...G...K...S...Q...A...P...O...C...P...L...E...T...R...P...L...-...R...E...T...R...K...C...I...S...Y...I...S...Y...E...M	179						
Db	722	GV...H...L...G...V...L...E...H...O...K...N...R...C...D...N...D...F...I...R...G...R...C...R...T...C...S...S...E...L...T...C...P...F...G...R...I...S...G...E...N...E...K...R...C...I...Y...T...S...Y...F...M	781						
Qy	180	K...R...V...T...G...C...O...P...C...V...R...I...T...T...R...A...C...W...L...A...S...I...A...H...N...A...K...A...P...E...G...V...K...M...A...L...G...T...A...S...V...M...D...G...V...M...G...T...G...C...Q...G	239						
Db	782	R...T...L...I...G...O...P...K...C...V...R...I...T...T...R...E...C...C...A...F...F...G...P...O...C...P...G...N...A...W...C...P...G...N...G...I...C...L...D...G...V...M...G...T...G...C...E...G	841						
Qy	240	L...G...F...N...T...A...C...E...T...E...G...K...Y...G...I...H...C...D...A...G...S...C...V...H...R...G...S...O...G...P...I...G...D...S...C...C...D...V...G...M...R...V...K...C...D...M...E...T...T...Y...N	299						
Db	842	B...F...S...F...G...A...C...E...T...E...G...K...Y...G...I...H...C...D...A...G...S...C...V...H...R...G...S...O...G...P...I...G...D...S...C...C...D...V...G...M...R...V...K...C...D...M...E...T...T...E...N	901						
Qy	300	C...N...G...T...H...T...S...A...N...C...L...I...D...D...G...K...A...S...C...K...C...A...G...F...R...G...N...T...V...T...C...A...I...N...A...C...T...S...N...G...G...S...T...R...A...D...C...K...R...T...T...P...N	359						
Db	902	C...N...G...T...H...T...S...A...N...C...L...I...D...D...G...K...A...S...C...K...C...A...G...F...O...Q...N...G...T...I...C...A...I...N...A...C...I...S...N...G...G...S...A...K...A...D...C...K...R...T...T...P...R	961						
Qy	360	R...V...C...V...K...A...G...Y...T...G...D...G...I...V...C...L...E...I...N...P...C...L...E...N...H...G...G...C...D...R...A...E...C...T...O...T...P...R...N...A...V...C...N...C...L...P...K...Y...T...G...D...G...V...C...L	419						
Db	962	R...V...C...T...K...A...G...Y...T...G...D...G...I...V...C...L...E...I...N...P...C...L...E...N...H...G...G...C...D...K...N...A...E...C...T...O...T...P...R...N...A...V...C...N...C...L...P...K...Y...T...G...D...G...V...C...L	1021						
Qy	420	I...N...V...C...L...T...N...N...G...G...C...P...F...A...F...C...V...N...T...E...D...O...R...I...C...T...C...P...Y...T...G...D...G...I...V...C...R...S...I...S...Y...E...L...P...K...N...T...S...O...Y...F...O	479						
Db	1022	I...N...V...C...L...T...N...G...G...C...E...P...F...A...I...C...H...N...T...G...O...V...E...R...T...C...T...C...K...P...Y...I...D...G...F...I...C...R...S...I...Y...O...E...L...P...K...N...T...S...O...Y...F...O	1081						
Qy	480	L...O...H...A...V...R...E...L...A...G...P...P...T...V...A...P...L...I...S...S...F...N...H...E...P...R...I...K...D...M...D...O...G...I...M...S...O...V...L...R...Y...H...V...G...C...O...L...L...D...L...N...K	539						
Db	1082	L...O...H...E...V...K...D...I...V...G...P...P...T...V...E...A...P...L...I...S...A...F...D...E...E...R...V...A...R...D...M...D...K...Y...G...I...M...P...O...V...L...R...Y...H...V...A...C...H...O...L...L...E...N...K	1141						
Qy	540	V...I...T...S...A...T...I...T...O...G...E...V...S...I...S...S...D...Y...I...F...I...N...N...E...A...V...L...S...D...I...I...S...T...N...G...I...H...V...I...D...I...K...L...S...P...N...L...I...T...P...R...D	599						
Db	1142	V...I...S...N...T...S...L...O...G...E...I...V...I...S...S...O...S...I...Y...I...N...K...K...I...I...S...D...I...I...S...T...N...G...I...H...I...D...I...K...L...S...P...N...L...I...T...P...R...D	1201						
Qy	600	A...L...G...R...V...I...O...N...T...T...V...A...N...H...G...Y...T...K...F...S...K...I...O...D...S...L...S...V...I...T...D...S...I...M...T...P...T...V...E...M...P...T...D...K...A...L...E...A...L...P...P...E...O	659						
Db	1202	N...S...G...R...I...O...N...T...T...I...A...T...N...N...G...Y...I...K...F...S...N...L...I...O...D...S...L...S...V...I...T...D...P...H...T...P...T...I...L...F...M...P...T...D...Q...A...L...H...A...L...P...A...B...O	1261						
Qy	660	D...F...L...N...O...D...N...K...D...K...L...S...Y...L...K...F...H...I...V...I...D...S...K...A...L...A...S...D...L...P...R...S...A...W...K...T...L...O...G...S...E...L...S...V...R...C...G...S...D...I...G...E...L	719						



```
Db 1262 DFLFNODNKKDKLEYLKFHVIRDAKVLAVDLPTSTAMKTLQSGSELSVYKCGAGRIDGLFTL 1321
Qy 720 NEOMGRFTHRGILLPDVGAAVYIDCLLMNPTLGGRCDDFTFTPDLPBEGSGCIFTTPCPLKS 779
Db 1332 NGQOTRIVORELLPDLGVAAVYIDCLLIDPTLGGRCDDFTFTPDLPBEGSGCIFTTPCPLKS 1381
Qy 780 KPKGVKKKCIYAPLPRFRNVEGCONLCTVYIOTPRCHGIFMPDQACGCGGDTFCNNNG 839
Db 1382 KPKGVKKKCIYAPLPRFRNVEGCONLCTVYIOTPRCHGIFMPDQACGCGGDTFCNNNG 1440
Qy 840 MCRDLPTMGGCLCTHGTNGTACELCMWGRFGPCDQPRSCSEHGOCDEGITSGECLCT 899
Db 1441 VCLDDYSATGECCKNTGNGTACELCMWGRFGPCDQPRSCSEHGOCDEGITSGECLCT 1500
Qy 900 GMTASCTPPAVAFVACVTPACSVHATCTENNTVCNLTNEGITICTVYDFECKONNGCA 959
Db 1501 GMTASCTPPAVAFVACVTPACSVHATCTENNTVCNLTNEGITICTVYDFECKONNGCA 1560
Qy 960 KYAKCSOKTOVSCSKGKGYKGDGYSCIEIDPCADGVNCGCHEHATCMTGPGKHCKECK 1019
Db 1561 KYAKCSOKTOVSCSKGKGYKGDGYSCIEIDPCADGVNCGCHEHATCMTGPGKHCKECK 1620
Qy 1020 SHYVGDVDCBEQOLPLDRCLDQNGCHPDASCADLYFODTTVGVFHLRSLPGQYKLPFD 1079
Db 1621 SHYVGDVDCBEQOLPLDRCLDQNGCHPDASCADLYFODTTVGVFHLRSLPGQYKLPFD 1680
Qy 1080 KAKKCAKEAATATYNNOLSTYOKAKYHLCGAGMLESGRVAAPTTYASOKCANVGYVD 1139
Db 1681 KAKKCAKEAATATYNNOLSTYOKAKYHLCGAGMLESGRVAAPTTYASOKCANVGYVD 1740
Qy 1140 YGSRAKSEMDVFCYRKMDVNCCTKAGYVGDGFSQGNLLOVLMSPSLTNFLTEVLAF 1199
Db 1741 YGSRAKSEMDVFCYRKMDVNCCTKAGYVGDGFSQGNLLOVLMSPSLTNFLTEVLAF 1800
Qy 1200 SKSSARGAFLKHLTDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNG 1259
Db 1801 SKSSARGAFLKHLTDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNG 1860
Qy 1260 TFLPMLISOLLITFPOSOHL-OETRFVYDGRSIILOMDITIAANGILHIISEPRAPTAT 1318
Db 1861 TFLPMLISOLLITFPOSOHL-OETRFVYDGRSIILOMDITIAANGILHIISEPRAPTAT 1920
Qy 1319 AAHSGLTGIFCAVVLVTGATLAALAAVSFRLKQRTTGFORFDOKRTLMWMLAS----- 1372
Db 1921 LHTGLGAGIFPAIILVTGAVALAASVFRINRRTIGFQHFSEEDINVAALGKQOPEMI 1980
Qy 1373 SSPRISOTLCMRPQRHROSPPVTPSQ 1399
Db 1981 SNPLYESTTSAPPE--PSYDPTDSE 2004
```

```
RESULT 3
US-10-104-047-2580
; Sequence 2580, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2580
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2580
```

Query Match 13.9%; Score 1094; DB 6; Length 510;  
Best Local Similarity 47.8%; Pred. No. 7.7e-54;

```
Matches 193; Conservative 60; Mismatches 149; Indels 2; Gaps 2;
Qy 849 GGCCLHTGTNGTACELCMWGRFGPCDQPRSCSEHGOCDEGITSGECLCTETGTATASCDPT 908
Db 8 GGCCLHTGTNGTACELCMWGRFGPCDQPRSCSEHGOCDEGITSGECLCTETGTATASCDPT 908
Qy 909 PTAFAVACVTPACSVHATCTENNTVCNLTNEGITICTVYDFECKONNGCAKYAKCSOKG 968
Db 909 PTAFAVACVTPACSVHATCTENNTVCNLTNEGITICTVYDFECKONNGCAKYAKCSOKG 968
Qy 969 TVSCSKKGYKGDGYSCIEIDPCADGVNCGCHEHATCMTGPGKHCKECKSHYVGDV 1028
Db 969 TVSCSKKGYKGDGYSCIEIDPCADGVNCGCHEHATCMTGPGKHCKECKSHYVGDV 1028
Qy 1029 C-EPEDLPDRCLDQNGCHPDASCADLYFODTTVGVFHLRSLPGQYKLPFDKAKCAK 1087
Db 1029 C-EPEDLPDRCLDQNGCHPDASCADLYFODTTVGVFHLRSLPGQYKLPFDKAKCAK 1087
Qy 1088 EAAITATYNNOLSTYOKAKYHLCGAGMLESGRVAAPTTYASOKCANVGYVDYGRANKS 1147
Db 1088 EAAITATYNNOLSTYOKAKYHLCGAGMLESGRVAAPTTYASOKCANVGYVDYGRANKS 1147
Qy 1148 EMDVFCYRKMDVNCCTKAGYVGDGFS-QSGNLLQVLMSPSLTNFLTEVLAFSKSSARG 1206
Db 1148 EMDVFCYRKMDVNCCTKAGYVGDGFS-QSGNLLQVLMSPSLTNFLTEVLAFSKSSARG 1206
Qy 1207 QAFKLHLDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNGTFLRTM 1250
Db 1207 QAFKLHLDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNGTFLRTM 1250
Qy 1260 TFLPMLISOLLITFPOSOHL-OETRFVYDGRSIILOMDITIAANGILHIISEPRAPTAT 1318
Db 1260 TFLPMLISOLLITFPOSOHL-OETRFVYDGRSIILOMDITIAANGILHIISEPRAPTAT 1318
```

```
RESULT 4
US-10-104-047-3794
; Sequence 3794, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3794
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3794
```

Query Match 9.7%; Score 762.5; DB 6; Length 482;  
Best Local Similarity 38.6%; Pred. No. 2.3e-35;  
Matches 164; Conservative 57; Mismatches 159; Indels 45; Gaps 8;

```
Qy 971 VSCSKKGYKGDGYSCIEIDPCADGVNCGCHEHATCMTGPGKHCKECKSHYVGDVDC- 1029
Db 971 VSCSKKGYKGDGYSCIEIDPCADGVNCGCHEHATCMTGPGKHCKECKSHYVGDVDC- 1029
Qy 1030 EPEQLPLDRCLDQNGCHPDASCADLYFODTTVGVFHLRSLPGQYKLPFDKAKCAK 1089
Db 1030 EPEQLPLDRCLDQNGCHPDASCADLYFODTTVGVFHLRSLPGQYKLPFDKAKCAK 1089
Qy 62 ESEPRVDRCLDQPRCHSDAMCTDLHFQEKRAGVNHLQATSSGYSGLNFSSEAEAC 121
Db 62 ESEPRVDRCLDQPRCHSDAMCTDLHFQEKRAGVNHLQATSSGYSGLNFSSEAEAC 121
Qy 1090 ATATATYNNOLSTYOKAKYHLCGAGMLESGRVAAPTTYASOKCANVGYVDYGRANKSEM 1149
Db 1090 ATATATYNNOLSTYOKAKYHLCGAGMLESGRVAAPTTYASOKCANVGYVDYGRANKSEM 1149
Qy 1150 WDFCYRKMDVNCCTKAGYVGDGFS-QSGNLLQVLMSPSLTNFLTEVLAFSKSSARG 1208
Db 1150 WDFCYRKMDVNCCTKAGYVGDGFS-QSGNLLQVLMSPSLTNFLTEVLAFSKSSARG 1208
Qy 182 WDAVCFRVDVACRCRNGVYGDGISTCNGKLLDVLATATANSFTFYGMLLGYANATRG 241
Db 182 WDAVCFRVDVACRCRNGVYGDGISTCNGKLLDVLATATANSFTFYGMLLGYANATRG 241
Qy 1209 FLKHLTDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNGTFLRTM 1268
Db 1209 FLKHLTDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNGTFLRTM 1268
```

```
Db 242 FLDELDELTYKTLFVFNVEGVNMTLSGPDSELSHASNALLSAN-ASQKLLPAHSGL 300
QY 1269 QLLIT-----FSODQLHETRFVDSRILOMDIITA 1298
Db 301 SLIISDAGPDNSMAPVYSLATVPFCWMPSPSLTVGLITLVLSTLGAIPGVYSR-IIVMDIMA 359
QY 1299 ANGLIHIISEPLRAP-----TAATPAHSGLCTGIFCAVAVLTGAI--ALAAVSFRLKQ 1351
Db 360 FNGIIHALASPLAPPOQAVLAPAPVAAVG---AVLAAGALLGLVAGALYLRAG 415
QY 1352 RTTGF 1356
Db 416 KPTGF 420

RESULT 5
US-60-360-039-5280
; Sequence 5280, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360, 039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5280
; LENGTH: 1111
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-60-360-039-5280

Query Match 6.1%; Score 480.5; DB 7; Length 1111;
Best Local Similarity 22.2%; Pred. No. 3,1e-19;
Matches 251; Conservative 98; Mismatches 341; Indels 441; Gaps 70;

QY 115 NVATDKGVY-HGLEKYLEI-----QKNRCN-----NDTIYRGECKGSGQAPCLETR 163
Db 43 HVCYTKYIVDDYELKVIHTVYVNDTEQCLNPLTGFQCTVERKQ--KASYORQL--VK 97
QY 164 PLREFPKCIYSIFMGKRSVFIGCOPGV--RTIIRACWLASLANAPAGEVMCA- 220
Db 98 KERIVKQCCDGYOTKHCLPDCNPPCKKKKCIERKGC-----ECDPYIGC-KYCAS 149
QY 221 -----LGTASVWDVNGT-----GTCCGGLGFNGTACE-TCTEGKYGIHCDQASC 265
Db 150 SCVSVGTWGLGSGKSCDCENGANCDELGTCTICTSGFGGERCEKPCPDNKGWGPCVASCPC 209
QY 266 VH-GRCSQPLDGDSCDDCVGM-----GVYCDMEITITDNCNGCHTSANCL 311
Db 210 QWGGKCKN-----EGKVCVSDGWSGGERCLNCKEGRFGAECKE-----CN-CQNGATC- 257
QY 312 LPDGRKASCK-----CAAGFRNGTV--CTAIN--ACETSNG----- 344
Db 258 DMTNGKCTCKSGYHGALCENECVSGFFSGCTQKCDCLNQNNDSSGECCKTIGMTGKHC 317
QY 345 --GCST--KADCKR--TTPG-----NRVCYKAGYTGDIIVCLEINPLEN 384
Db 318 DIGCSRGFRGLQCKONCTPGLEFSDSNASCDAKTGCCCESGYKGP--KCDERKCDARQ 375
QY 385 HGC-CRNAAECTY-----TGPNAVCNCLPKYITGDKVSLINVLCTNNG--CSPFARC 436
Db 376 YGADCKSKTCTCAVRENTLMCAPINTGFCRCKPGEYGDN--CEL--ACSKDSYGPNCERQANC 431
QY 437 NTEDOQ-----RICTCKPDYTGDIIVCGSIYGLPKNPSTSQYFQLOEHAVRELACG 491
Db 432 DMNHASECNPETGSCVCKRFGRTG-----KNCSE-----P 460
```

```
QY 492 GPFTVFAPLSSSFNNEPRKMDDOGLMSOVLRIHVCGCOQLLDNLKATTSATTLQGP 551
Db 461 CPLEDYGP--NCAHQCC--NORG-----VGD----- 484
QY 552 VSISSQDTVFINNBAKVLSSDIISTNGVIVHIDLKLSPKNLLITPKDALGRVLONLTV 611
Db 485 -----GADCKCCDDGWGTH 499
QY 612 AANHGYTKFSKLIODSGLSVITDSIHPPVTFVWPTDKALEALPPEODFLPNODNKDL 671
Db 500 RCEH-----HCPADTF----- 510
QY 672 KSYLKFHVIRDSKALASDLPRASWKTLOGSELSVRCYSGSDIG--ELFLNQMCRFHNG 730
Db 511 -----GANCKRCKCPKIGIDCDPTGECTCPAGLQ 541
QY 731 LLEFVGV--AYGIDCLMNPILGRCDFTFTFDIPGEGSC---ITFPKPKLK-SKPKG 783
Db 542 ANCDIGCEPGSYGPGCKLHCKCYNGCKOKET-----GEC-TCQPGPFSSDCSTTSKGR- 594
QY 784 YKKKCIYNELPFRNVEGQNLG-----TVVIQTPEPCGHYFMPDQACPGGP-----DT 833
Db 595 -----YG-----ESCELSGSDASCSKOTGK-----LCPLGTKGVSQDQ 630
QY 834 PC--NNRG-MCRDLITPM-----GCCL-CHTFGNFTACE-LQWHRFGPDCOP-RS 878
Db 631 KCDPNTFGLCOETYPSPCASTDPKNGVCLSPRESSGHEHNCPASYSDDGCOQVCS 690
QY 879 CSEHGQCDGITGSGECLEGTGTAAACD--TPTAVF-----AVCTPACSVHATCTE-NTT 931
Db 691 CADHGCD--PTTGECICEPBYHGKTCSEKCPDGKYGIGCALDCKKCKAGSGTCHINL 747
QY 932 CYCNLNEBGDITGVVDFCKONNGCAKAVAKCSOKGTVS-----CSCKKYGKDGYSK 986
Db 748 CIOPAGE--GALCTRPSCSAGFMGNCROVCRCTSEYKQCMQOTGSCPPAGFGOD--RC 803
QY 987 IETDPCADGVN-GCHEATCKMTGPG-----KHKCEKSHVGD-----GVNC 1029
Db 804 DK--PCDEYGPDCIKKCKCGGTATSSCNRVSGACHCHPGTGECHALCESTFFGLKC 861
QY 1030 EPEQLPLDRCLDNGCHPDASCADLYFODTIVGVFHL-RSPLGQYKLTFD 1079
Db 862 SKF-----C-PKDCGGDYECDALIGCHVQMSGKRAKOFEE 898

RESULT 6
US-60-360-039-5784
; Sequence 5784, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360, 039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5784
; LENGTH: 3507
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-60-360-039-5784

Query Match 5.1%; Score 399; DB 7; Length 3507;
Best Local Similarity 16.7%; Pred. No. 3,8e-14;
Matches 316; Conservative 153; Mismatches 490; Indels 932; Gaps 74;

QY 134 KNRCDNNDTIYVGECKGCSQAAPCLET-KPLRETRKCIYSIFMGKRSVFIGCOP--Q 190
```

Db 1537 KKPDSTOS--SKSHCSNMS--CEVDTVDGSECKEOMGKKSGK-----VCEDINE 1587  
QY 191 CVRFTITRACMLASLAHNAKPAPGEVKKMCAIGTASVMDGVNGTGCOCGLGPNG----- 244  
Db 1588 CV-----AEKAP-----CSLANCV--NMNGTFCSCOGYRGDGFMCFT 1624  
QY 245 --TACE-----TCF--EGKYGIHC-----DOAC----- 263  
Db 1625 DINEDEHHPCHPAHECTNLBESSFCBECHSGFEQGIKCTPTPLERSGCDVEKFCGRVDH 1684  
QY 264 -SCVGRCSOGPLGDGSCDDCVGMRGV--CDMEITDNC--NGTCHTSANCLDLPD 315  
Db 1685 VSLCLVRIYNGSL--SSVCECEGFPEKESNSC--VIDECEBSRNMCDPASAVCVNTE 1740  
QY 316 GKASCKACAGFRNGCTVCTAIINACTSNGSGCTKADCKRTTPTGNVAVCYCKAGTYDGIYC 375  
Db 1741 GSYRCECAEGTEGEGVCTDIDECRGMAGCDSMACINRM--GSCCCCKMAGYTDGATC 1799  
QY 376 LEI-----NPLENH----- 385  
Db 1800 KIEEPEKSDKTACTDEMSRLCELEKKOCTVDEEEVPOCGACLPGHHPINGTCSQSLQISG 1859  
QY 386 ----GGCDRNAECTGTGPRNOAVCNCLPKRYTGDGVCSLINVCLTNNGCGSPAFACN--Y 438  
Db 1860 LCAOKNDCKNHAECIDIPDISHFCSPDGFIDGMIICDDVDEC--NAG-----MCDDEN 1912  
QY 439 TEODORI--CTC-----KPDYTC- 454  
Db 1913 TKCENTIOSFNCVCLLEGKKKDEKCVDEKKOPNREKIEIDENSSSSNGOEKPTTKGI 1972  
QY 455 ----- 454  
Db 1973 VSSTSATSESTTAPHVHTSISSTTKDMTSSKSPENVMTSSSEPVSTSSSKSTTAS 2032  
QY 455 -----DGIVCRGSITGELPKNPSTQYFPOLQ-----E 482  
Db 2033 ETTVSTPSESSSEAPLTSSPATTEVTETSESVKSTPKESSESEITVKLSKSPVEYE 2092  
QY 483 HAVRE-----LAGPFPYFAP-----LSSFN- 505  
Db 2093 SSVASSPSTTSSQSVTSTVPEITSKSTVLSSEAPVSTSPREVTSETPKPSLASSSTT 2152  
QY 506 -----HEPRIKMD-----OOG 517  
Db 2153 GDTNSTPTSTSLASVKSSTAPEGTSASVAPYKLSLSPDVSOPTKTFDAPESSTVQAS 2212  
QY 518 LMSOVLRIHVHVCQOOLLNDNLKVTTS-----ATTLOGEPU 552  
Db 2213 ETSSGTSVKSSTSEPSHVTKLSTSSNPSSSVPTSPKSTPTVPESTEQPTSTPSSQSL 2272  
QY 553 SISVSODTVFINNEKAVLSS----- 572  
Db 2273 TPMNSNSEVLTSTSEPHVLSSLSLSPDVSSQSTTPNNLSESTVETPKTSEVSLNSEEPST 2332  
QY 573 -----DIIS-----TNGVIHVIDKL 588  
Db 2333 TEAPPTLSPLDLSLTNNLSOSSIVSTEDRSFISENSEKPTSAPELVTSSVTHVAASSP 2392  
QY 589 SPKNLITPKALGRVNL-----TTVAANH----- 615  
Db 2393 DVPTESSEPDLTGSTEINPEASSKOTISSTPPDTTTASSEPKSTSMSPDLSTTSNV 2452  
QY 616 -----GYTKFSKLQODSGLSVITDSI--HTPVVFMPT 647  
Db 2453 LSESTTPSESSKSPVSSSTEGISVYSTEFSK--VESTISSVLEEDLTKTTPSPILEBT 2511  
QY 648 DKALEALPPEODFL----- 662  
Db 2512 TTASPTSPLEEDSLTVSVRIHELTTSENVPKSESTTTSSSESSKSPQEPAGILITTV 2571  
QY 663 -----FNODN-----KDKLSYIKFHVIR----- 681  
Db 2572 VPTSSVSLITASEIEAITSNTPFKQGRPTTSPKSLVKSTTSPSTVSSSESTKRTT 2631

QY 682 ----- 681  
Db 2632 VSTTVSTTTPTTETTTSSLLITTAAPSKPTTESTESSEAPTTTPATSETKPSNVSTSK 2691  
QY 682 -----DSKALASDLPKASMTLOGSELS 705  
Db 2692 STENETSTNSQSLSESSMSTSSPETNAPAVYSSASSSTLTLENNSTSSPTSSSELS 2751  
QY 706 VACGTGSDIGELFLNE-----OMCRFIRGLLF-----DV 735  
Db 2752 YK--LSSLFESITSEAVTVSSRAPELITMSSESHREISTVSSSEPEEIPLTSTVSPV 2809  
QY 736 GVAYGI-----DCLLMN--PTLGRCDFTTFDI.PGBCG 767  
Db 2810 VTASSIPSEEPILSVTSSSTPRVRLTGTPODLIVSYTVPSHGNNRONTASSVPSNST 2869  
QY 768 SCIF-----TPKCP--LKSXPKGVKK-----CIYNPLP-----FR 796  
Db 2870 SPILPSESLTTPQPPPTTTAKPATTSGRKGPSIQPPAEMFTTPAPPPSNGCYGE 2929  
QY 797 RNEBCCQNLCTVVIOTPRCHGYFMPDQACPGPDTPCNNRGMKRDLYTPMGQCLHTG 856  
Db 2930 TNOEEQVYSTTTTEAPSLC-----STVYCHSLATGEQ--STGVICICRDG 2972  
QY 857 ENG--TACELCWHGRFGPDC--OPRSCSEHGCGDEGITGSECLCEIQTWTASCDTPTA 911  
Db 2973 FTGDGTTACS--KSTADCLSLPSLCAADKACD--NSTSCBCDAGYIGDG----- 3019  
QY 912 VFACVP-----ACSVHATCTENNTVCNLTNEBQDITCTVAVVPCKONNGCAKV 961  
Db 3020 --YVCSPPHQDCVLKNDLNSPEAVC-QNRRCOCLPFGTGDGVKCVSIHERANCSQCDAN 3076  
QY 962 AKCSQKGTQVSCSKKKGKGDGYSCEIETDP--CADGVNGCHEATCRMTGPKKRCCK 1019  
Db 3077 AHC-VGCT--TCKCNPBGYGNGLCCVP--DPLDCVH-FTGICHPNAYCNSE--SHQCCS 3128  
QY 1020 SHYVGDVDCPEPEOLPDLRCLODNGCQHPDASCADLYRQDTYGVFHLRSPIGQ----- 1073  
Db 3129 SEFSNGVSCFPOK--SCRTDKSVCAKMAICL-----PGSCICIRMG 3168  
QY 1074 YK-LTFEAKAKEACAEATITATYNOLSYAQ--KAKYHLSAG--WLESGRAVYPTTAS 1127  
Db 3169 FKGDPPFYKCSLVSXKPEPANQODLSDVSSCVTPPCDSSTQCLISGECICKSGFRNRTLSGS 3228  
QY 1128 QKCGANVVGIVDYSBRANKSEMDVFCYRMKDVN-----CTKAGYVGDFSCSGNL 1180  
Db 3229 ETC-----ADIDCAEKSHK-----CDRVATCRNTFGSHVCTCPDGHVGDGITCVPHVN 3277  
QY 1181 QVLMSEFSLTNTLEVLIAFSSKSSAGQAFK 1211  
Db 3278 QGRLSVYCCBADGMTLVLGNETADFEGRKIFVK 3308  
  
RESULT 7  
US-60-360-039-6549  
; Sequence 6549, Application US/60360039  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Chen, Xianfeng  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)A  
; CURRENT APPLICATION NUMBER: US/60/360,039  
; CURRENT FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6349  
; LENGTH: 1584  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans

US-60-360-039-6549

Query Match	4.88;	Score 375.5;	DB 7;	Length 1584;
Best Local Similarity	20.28;	Pred. NO. 3.3e-13;		
Matches 242;	Conservative 125;	Mismatches 384;	Indels 449;	Gaps 65

QY	18	FRGTYIHVNLASLESADATVAEPNNEA1-----EWYIEKK---ATSLKEJDLRYHV	69
Db	502	FSGNHIDYFRSHLFVSGTLPDIPARSEVLFPEDEEYVERGGLYLSKAAFDV1-----	556
QY	70	LGEKLNLDLHNGHRETMGLGFSYLLAFELRNDOLYVNEAP1VNTVATDGVIGHLEKY	129
Db	557	-----VADGQNVOT-----YRMSV-----DQITFEBCP-----NKEFRD--HSMK--	591
QY	130	LEIOK-NRCNNND1IYRGECKGKSQAOPPLEKPL--RETAKCIYSIFMCKRSVYFG	186
Db	592	LHVRIINYVDDGGVYRGAKNKATFATVSPRAVASPFGHFDHRROHQSMMASERPLEIP	651
QY	187	COPCQVFTIITRACWMLASLHNAKPARGEVKMCALGTASVWDVNGTGCOCGLGFN---	243
Db	652	SQSSIS1F-----DSVCARGR-HQCTLPRMKC-RVYDPSYRECEPBYQAH	696
QY	244	-----GTACTCTEKGYYG1H-CDQASCYVHRCSSQGPLDDGSCDDCVGR--GVKCMDE	294
Db	697	DASSHIGMICODLDECORGHNCDO-----HAKTNAP-GSFSCOC1OGYQSGRRC1RE	750
QY	295	-----ITTD--NCNGTGTANCLL--DPDGKASCACAGFR	328
Db	751	HOASHHEHNOOTPEBMAGVGTITGFGCTAHNOCHEMGCEYFTSEHPLGR--CCKRGMYG	808
QY	329	NGTVCTAINACE-----TSNGG-----CSTKADCKRTTPGNRVACKAGY	368
Db	809	DG-----VNEHGPREENMPKHAN1PORGQACGSYVCADNACMEPSSGSECVKAGF	863
QY	369	TGDDIVYLE1EINPCLENNHGDDRMACEOTOPRNOAVCMCLPKYTGDKGVCGLIIVCLNNG	428
Db	864	SGNVCTESLDDRHAHSHHRRH0Q0TG-----SLGVCASHBEC-SEHG	909
QY	429	GCSPFACNTEODORICTCKPRDTGDIYCRS1YELPRKPNSTSOYFPOLOEHAVREL	488
Db	910	SCAVHNSLGYYQ-----CACTEPYVNGVBC-----TLRGSSAS-----	943
QY	489	AGRPFIYFAFLSSPFNHEPRIKIMDQOGLMSOVLRYHVVCQOLLN1KYTTSATTLQ	548
Db	944	-----AP-----ELGCD-----VTRDC-----	955
QY	549	GEPIYSIVSOPTF1INNEAKVLSDDL1STNGVIHVIDKLSPKNLITPKDALGRLVNL	608
Db	956	-----SEPDACVYERS-----STATFRCV-----	975
QY	609	TTVAANHGYTFKSLDIDSG1-LSVITD1SHTPYTVFWPTDKALEALPREQODFLENODN	667
Db	976	-----CQSGYTCDDGYCKWQSOLAI1SPSV-----POLPSEP-----	1000
QY	668	KDKLSY1LKFHVIRDSKALASDLPRASAMTKLOOSELSVEGCGSPD1GELFLEDMOCRF1	727
Db	1009	-----AVL1AS1CNPMCGGDAOCVYDHNK0-----	103
QY	728	HRLGLFDVGAUYIDCL-LMNPTLGRCDTFTTFDIPREGCSOFTFKPKCLSKRPVYK	786
Db	1035	YRCCEYAAFMGDGNCVPLAKPNN-----YPAQKTCVESDCH1NG-----	107
QY	787	KCIYNP1LPRFRN1VGGCQNL1VY1QTRCGHGYFMPCQACPRG-----GP	831
Db	1077	HCV1NE-----HGAGEY1CQCL1PEFSGDGF1NCRGADQCN	111
QY	832	DTP--CNNRKCRC-RDLYTPMGOCLCHTGF--NGTACELCWHGRFGP-----DQPRCSSEH	882
Db	1113	SNP5AC1QYONAC1VYDAL1LNHAC1CVGDFRGDGTSCV-----PYAPAT1CN1LEPR1CHAN	116
QY	883	GQC-----DEG1TSGGEC1LCTGWT1A--ASCD--1PTAVFAVCTPAC1SVHATCTENNT--	931

```

Db 1168 AQCVMHDD---TAAVEICIKPSSGGGYTKCDVIEPR---CT-NCSIHAYCAQNPST 1218
QY 932 ----CVCNLNTEEDDGIITCVNDFCKONNGGCAKVAKC--SQKGTQVSCSCCKKTKGDGYS 985
Db 1219 GAYQCKCNAGATYNGNGLHCYSMSSCLDRLSDENADCVPEAGGHYV--CNCHGYGHGDGRS 1277
QY 986 CI-EIDPCAGGVGCGCHEHAHATCTMTPG---GKHKECKSHY--VGGGVQCEPEQLD--- 1037
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1278 CSPSSSTRSKLLVARKMALFERSTNPDEYQGLIVPHILIPVGIDFCKEKEKIWSDMS 1337
QY 1038 ----RCGLDDNGQCHP-----DASCDALYQDPF---TVGPFHLRSPILGY 1075
Db 1338 GHSTRISLNGTEKSTYFNKELSSPGCIADVDMSSRNVIYADSMDEIGASLN---GKFK 1394

```

## RESULT 8

```

US-60-360-039-5709
: Sequence 5709, Application US/60360039
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xianfeng
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
: FILE REFERENCE: 38-10(52052)A
: CURRENT APPLICATION NUMBER: US/60/360,039
: CURRENT FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 5709
: LENGTH: 1106
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
US-60-360-039-5709

```

Query Match 4.8%; Score 373.5; DB 7; Length 1106;  
Best Local Similarity 19.3%; Pred. No. 2.9e-13;  
Matches 193; Conservative 85; Mismatches 265; Indels 459; Gaps 42.

[illegible]

```

OY 532 QLLDNLKKTTSTATLQGEFVSVISVQDYLFINNEAVLSDDISTNCVHIVDKLSPK 591
Db 588 -----SSTAA---PI-----LSEITF---GKRYSKYIVSTNGII----- 616
OY 592 NLLITPKDALGRVLQNLTTVAANHGYKFKFSKLIQDSGLLSVITDSIHTPTVFWFPTKAL 651
Db 617 -----SFDGDLQGL-----IDHAETLKKRAIFAL 640
OY 652 EALPPEQDLEFNQDNKKDKLSYLFKHVINDSKLASDLPRASAWKTLQGSSELSVRCGTG 711
Db 641 HA---QFDYI-----REGLVAYT---FINDTDVATPLP-----LMRSSIGIQTAMG 680
OY 712 SDIGLEFLNEOMCRFTIHRGLFPDGVAVYIDCLLMNPTLIGRCPTTFPFDIPGEGCS--- 768
Db 681 I---ENFQTKRLHIF-----TFDRVROSGSENF 705
OY 769 -----CIETPKCPKSKPKGVAKKCIYNLPERRNVEGCONLCTVYO 811
Db 706 NSFQIVLAESDKDATILSLIYERKVGARGPMTGISPSRFLELPPNNRLASGSN-----VGQ 760
OY 812 TTRCGHGATPMPDQACPGGDTPCNNRGMCRDLTTPMGQCLCHTGFNGTACELCHMGARG 871
Db 761 PKGMHTRVDIADLQOTCPG-----RLGEPLCD-----RECAAHNG 796
OY 872 PDCQPR-SCSEHGQCEGLITSGEC---LCETGWTAAACDPTPAVFAVCTPACSVHATCT 927
Db 797 INCESTCHCDGSVACDV-ITGM--CGPGLCRAGWEGSSCD-----Q 834
OY 928 ENNTCYCNLYNEEDDITICTIYVDFCKONNGCCAKYAKCSQKCTOVSQSCCKKGYKGDGYSCI 987
Db 835 DIDECEMSTL-----YTCAGSQCQVNTRGG-----YRCDCGFAPVYKECK 875
OY 988 EIDPCADGVNGGCHENATCRMTGTGKHKCECKSHYVGDGYDC 1029
Db 876 PIDRCLSRFSVPCSRNAECYESTIESNPKCYCKRGYHGDGFC 917

RESULT 9
; Sequence 5866, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360, 039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5866
; LENGTH: 3051
; TYPE: prt
; ORGANISM: Caenorhabditis elegans
US-60-360-039-5866

```

```

Query Match          4.7%: Score 371; DB 7; Length 3051;
Best Local Similarity 20.4%: Pred. No. 1.2e-12;
Matches 244; Conservative 109; Mismatches 427; Indels 416; Gaps 68;

QY 133 QKNRCDNNT-----IIVGEC---GKCSQARCPLETKPLRTRCIISTYFM 178
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 818 EMERCTNDGSDWFCCELPQFERIRMGOCAPGSCNPNDEMSCD---VRRKROCL----- 868

QY 179 GKRSVFEGQ-----PQVRF--ITTRACWLASLAH 207
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 869 -PRGNIIYTCQGRNEKRHPETDICKNEQTGEHDCDRSARCIDPDESYICACQSGFIDH 927

QY 208 NAKPAPGEKKMALGTASVMDGVNLTG-----TCQCGIG-----FNCT 245
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Dd	928	SPNSERGRVCAVLAONECLDGSNRCSPNALCTDTEEGYVCRCKSSFDVYSPNPOTFFGM	987
Qy	246	ACE-----TCEGKXGIIHCDOACSCVHGRCSCQPLAGDSCDCDGVW-----RGVKCDM	293
Dd	988	VCKELVEBECTNPRILN--QCDRANHCI-----DTEBYSOICRPGFVDMDFGNPCRBEO	1040
Qy	294	EITTDNCRN--GTCHISANCLLPDQKASCKAAGRGN-----GVYC-----T	334
Dd	1041	IKTMDKCSPGKNCDDRNARNCIQIDDDYSACAPRPGFKDSSPSRGRGLCPVIECPDN	1100
Qy	335	AINMCEISNGGCSFKKADCKRTTPGNHRCVCKAY-----TGQIVLEL--NCLCNH	385
Dd	1101	TLNDCSDP-----RAYCTDIDG--YMCRCQGFLDISPSISVKEPRLCKPLQNECALGI	1154
Qy	386	GGCRRNAECTQGTGNOAVNCLPKY-----TGQKVCSS--LINVLTLNNGCSPFAFC	436
Dd	1155	DDCARDDGICGDNDSFTCKCANNYLDVSDRYTRGRCKRCKRLINCCQYQNDGSELEATC	1214
Qy	437	NYTEDDORICTC-----KPDYTG--DGIVCRGSYIGELLPKNSTSYQFFQLOEHAUREL	488
Dd	1215	TDTE--DSYICACPOSHIDLSPTDTPNRRGRCLMRI-----NECSIN-----RHD	1257
Qy	489	AGCPGFYFAFLSSFFNHERID-----WDQGLMSOVLRYHVYCCQQLLDMLMKYTT	542
Dd	1258	CSPNADCTDPES---YKCRKHDFVDESDSSRRPRGJ-----CRLPVLDECRGK	1306
Qy	543	S-----ATTLOGEPAVISVSODTVEINNEAKVILSSDIITNGVILHVIDLKLSKNLLIPK	598
Dd	1307	HDCIVNALCQDLPGYTCQCSADFPV-----VSHRASHR--GRICQPRTPRPPE	1355
Qy	599	---DALGRVLONLTTVAANH-----YKFSKLIQDSGLLSVITDSTHTPVTYFW	645
Dd	1356	CRLDGNGQCKHLENYCRLMGEPKSCSPVNYORDS-----SGSCSLINECLFT-----	1404
Qy	646	PTDKALELAPREODPLFNDQNMKDKLSYLAKFHVITDSALASDLPRSMKMTLOGSELS	705
Dd	1405	-----QLNDCHTAACTIDQYQVY-----T	1423
Qy	706	VRCGTG--SDIGEL-----FLNEQMRFIH-----RGLLFVGYAYGIDCL--LMN	747
Dd	1424	CQCRDGEFIDIDRRRPRRCKPWAYNE--COYPIILNCHQNAACIDLEBETBECKMOGFM	1481
Qy	748	PTLIG--GR--CDFTTTPDIPEGSCSIFTPKCPKLSKPKGVKKKC-----IYNPLD-----	794
Dd	1482	HSHGPRGRIKQLTNNECLRPLSLNSCDRNARC--IDKEEGYEECBRGFIDVSPPTLGR	1539
Qy	795	--FRRAVECCONL-----CTVYIQTPRCCHGYFMPDOCAP-----GGDTP-----	834
Dd	1540	ACRELVEBCASRLNDCD--KNARKCKDTMDSYECDC--CEVNSKDISPSPSPRGVCLMET	1595
Qy	835	-----CNRRCMGCD--LYTPMGQCLCHIGF-----NGTAC-----ELCMHG	868
Dd	1596	NECSGVHDDCPSATYCKDNDSF-----CECPGFYDRSPNKHARPGVAYLVDECBEG	1651
Qy	869	RFGPCDQPRSCSEHQOQ--DEGTTGSGBELCTGTWTAAS-----CDTPRAYAV	915
Dd	1652	R-----HTCSSHADRDRLDEBGT--CECDMDGVDSPNLASDPGRVCAP-----EV	1696
Qy	916	CTP--ACSVHATCTE---NNTCVCLNLEBGDITCTVVPFCOON-----GGCAVYA	962
Dd	1697	CPRHNDSSAAVCEPRLGMRKYOCVCIQGYVDSQPSQKGRVCVRRNACHDRPLNFCSSNA	1756
Qy	963	KCSQKGTQVSSCKEKG-----KGGVYSC-----IELDPCADCVNGGCHAHATR	1007
Dd	1757	ICYDEPRGYRCECKRGEMDRSPDSSQGRKYCEPRPPSPRRHPCODPERRNDCHPAGTGR	1816
Qy	1008	MTGEGKHCECKSHYVGDGVDECEPOLPLDRCL-----ODNGCHPDASCADL	1055
Dd	1817	ATGAGOSTYCECLSYADRSPD--PRANKPGLVLTPEPCLADPBQNDCHAAALCSEV	1870



[illegible]

```

RESULT 12
PCT-US02-09671--703
; Sequence 703, Application PC/TUS0209671
; GENERAL INFORMATION:
; APPLICANT: Zyccos Inc.
; TITLE OF INVENTION: TRANSLATIONAL PROFILING
; FILE REFERENCE: 08191-026W01
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/336,780
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 703
; LENGTH: 3494
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671--703

```

Query Match	4.3%	Score 338	DB 1	Length 3494
Best Local Similarity	31.4%	Pred. No. 9.6e-11		
Matches 101	Conservative 30	Mismatches 125	Indels 66	Gaps 21
QY	182	SVFICGCPQCVRTIITTRACWLASLHNNAKPAGEKMCALGTASVWDVNGTGTCCGCG	241	
DB	24	SFFCICPPQ-----WKGPL-----CADVNECHITYSGTPLSCONG--GTCVNTMG	66	
QY	242	ENGTACETCTECKYGIHC-----DQACSCVHRC-----SQQPLDGSCDDCVGN-	286	
DB	67	--SYSCH--CPPEYGPQCAKYDDEEGGSVACVCHGICEDLMREQAGEPKYSCVCDAGM	123	
QY	287	---RCVKCOMETTTNCGN---GTCHTSAACLLDPPQKASC--KCAAGFRNGRVCAINAC	339	
DB	124	SSPNSPACTLID--RDCCSQPPGPCSTLYOC--FNTGSPFYCGACPTGWDNGYICEDINFC	180	
QY	340	ETSNNGCGST--KADCKRTTPGNRV--VCKAGTGGGIVLEINPLLENHGGCDRNAECTQ	396	

Db 181 EINNCGGCVAPPEVCV -NTPESSHCQACPPEYQDGRVCLITDLCYSNCGCHPDASCSS 239

Qy 397 TGPNAVNCCLPKYTGDS---KYCSLINWCLJN---NGGCSPAFCNPTEDORICTCK 449

Db 240 TLGSLPLCTCLPYGTGNGNGYGPNGCVQLSNICLSHPCLNGCID-TVSGY-----FKCD 292

Qy 450 PDYTGDIYCRGSIYGEI,PKKP 471

Db 293 SGWT--GVNCTENI-NECLSNP 311

```

RESULT 13
US-10-150-821-4
; Sequence 4, Application US/10150821
; GENERAL INFORMATION:
; APPLICANT: Welcher, Andrew A.
; APPLICANT: Elliott, Gary S.
; TITLE OF INVENTION: C3B/C4B COMPLEMENT RECEPTOR-LIKE MOLECULES AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 01017/37592
; CURRENT APPLICATION NUMBER: US/10/150,821
; CURRENT FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: US/09/911,842
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: US 60/222,438
; PRIOR FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 3594
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1757)..()
; OTHER INFORMATION: Xaa = any or unknown amino acid
US-10-150-821-4

```

Query Match	4.3%	Score 338	DB 6	Length 3594
Best Local Similarity	20.8%	Pred. No. 9.9e-11		
Matches 336	Conservative 156	Mismatches 555	Indels 570	Gaps 98
137	CNNDFITIVRGCGKCSOAPCPPL--ETKPLRETRKCIYSITYFMGKRREVFICGCPQCYR	193		
Db	1115	CPETTTTVYKRGAVDISACCVPCPVEBFSISGLITPCYPCRDYGYQPNACKSFCFLACPFGYT	1174	
QY	194	TTTTRA-----CWLASLAHNAKP-----APG-----EVKM-----CALGTASVMDGV	230	
Db	1175	TTTTGATATSLTDCSSFSPTSAEESIIVPLVAGHSUNKYEVSQVHECFLLMPC-----	1228	
QY	231	NETGTCQ-----CGLGFNGTACNETBECKYGIHCOQAS--CVHGRCSDGPIGDGS	279	
Db	1229	INHSGTCCQGLRGRTVCLCPRTGYGLKCT-----DIDECSSLPLCLNIGCRDQVGCTT	1280	
QY	280	CCDQDVGWRGVKCDMEITTDNC--NGTCHTSANCLLDPDGKAS--CKCAAFRGNGTVC--TA	335	
Db	1281	CCSGSLGYSGQICEENI--NECISSPCLNKGTC---TDGLASRYCVCVAGYM--GVHCETD	1333	
QY	336	INACGSENSGGCTKADCKRTTTPGNRVYCVKAKAYTDDGIYC--LEINPCLLENHGGCPRNAEC	394	
Db	1334	VNECOSS--PCLNNNAVCKDQV--GFGFSCKPQFLGTR--CEKNVDECSLQ--PCONGATC	1386	
QY	395	NOTGPNQAVNCNLPRYTGDGKVCSE--INVCLINNGGCSFPARCNTEODORICTCKPDYT	453	
Db	1387	KDGNANSFPCQCPAGPTGTH--CELININCGSN--PCRNQATC--VDELINYSCKQCPGFS	1440	
QY	454	GGGIYCRGSIYELPRKNPSTSYFPOLOEHAVRELAGPQPTVFYAPLASSFNHPRIKDM	513	
Db	1441	G-----HRCETDQ-----PSGFN-----LDF	1456	
QY	514	DOOGILMSQVLRHYHVGCOOLLDNLKVTTSATLL-----OGPEVTSISYSDG---	559	

Db 1457 EVSGIYGV-----LLDGLPTLHAIVCAFMKSSDVINYGPISIALEDKDN 1505  
Qy 560 -----TVPINNEKAYISDIIISNGYIH-----VIDKILSPKLL 594  
Db 1506 TELLIDYNGWLYVNGEKEKTIKPSVN-DGIWHIAITWISIGCAMRYIDIGELSDGDTG 1564  
Qy 595 ITPKDAL---GRVLONLTVAANHGYTKFSKLIODSGLSVITDSIHPTVTFMPTKAL 651  
Db 1565 LSIKKAIPGGALVLDGEODKKEGFPNPAESFVGSISGLN-LMDYVLSPPQV-----KLL 1618  
Qy 652 EALPPEQ-----ODEFLNODNKDKLSYLFVHIRSKALASDLP--RSASWKLQ 700  
Db 1619 ASSCPPELSRGNVLAMPDFLSGITGKVKVSSSMF--CSDPPSLEGSVPHLRPASGNKP 1676  
Qy 701 GSELVRCGTGSD-IG---ELFLNEQM-----CRFIHRGLL-----FDVGV 737  
Db 1677 GSKVSLFCDFQFQWVGNVYCLNQGOWTOPLPHCERIRGLPALENGFTSAEDFHAQS 1736  
Qy 738 AYCIDCLLMPPTLGR---CDTFTTPIRECGSCIFTPKCPKLS-----KPKGVK 785  
Db 1737 TVTYQCTSGYLLDSDSRMFCXDNQSMN--GISPCLDVBDECAVGSDEHASCINTNG-S 1793  
Qy 786 KCIILYNLPLFRNVEGCONLCVYIQTPRCHGYFMPDQACPGGPDTPCANNRGMCRDLY 845  
Db 1794 YVCSCNP-PYTGDKKNCAE-----PVKCK-----APENDENGSR-SGEIV 1831  
Qy 846 TPMQOCL---CHTGFN--GTACELC-----WHGRFPDCCORSCS-----FHG----- 883  
Db 1832 T-VGTAVTFSCDEGHELIVGVTITCLETGER-DRLRPSCEAISCAGVPVPEVNGVDGSAF 1889  
Qy 884 -----QCDEGITGSGE---CLCETGWTASCDPTTAVFAVCTPACSVAHATCENNT 931  
Db 1890 TYGSKVYVRCCKGYTLGSDDESACIASGWSHSS-----PVCGI-VKCSQGED 1936  
Qy 932 CVCMLNTEGOGITCTVDFCQNNNGGCAKAKCSQKGTQVSCSKKKGKGYSCIEIDP 991  
Db 1937 -INNGKTYLSGLTYLST-----ASYSCENGYSLQPSLLE--- 1970  
Qy 992 CADGVNGCHEHA--TCRMTGPRGKHC-----ECKSHYVGQDV-- 1028  
Db 1971 C---TAGSWMRARPSCOLVSCGEPYIKDAVITGNSFTFNTVAITCKEETIAGPRTI 2027  
Qy 1029 -CEPEQPLDRLCLDNGCCHPDASCADLYFODITVGVFHLRSLPGQYKLTFDKAEACAK 1087  
Db 2028 ICQAN---GKMNSNHQCLA-VSC-----DEPPNVDH-ASPETAHRLFDTAFFUYCA- 2074  
Qy 1088 EMATITATYNOLSTYOKAKYHICSA--GWLBSGRAVPTTYA--SOKGCAVNGIYDYSR 1143  
Db 2075 DGYSLADNSOL-----ICNMQGNVPPAGAVPRCIAHFCCKPPSVSYSLIESYK 2125  
Qy 1144 ANKSEMDVFCYRMKDVNCTKAGYV-----GDGFSCS-----GNLQ 1181  
Db 2126 A-KRAAGSVYSFK-----CMGEFVLNTSAKIEICLRGEMSPSLVQCIPIVRCGEPPS 2177  
Qy 1182 VLMSFPLJNE-LNEVLAFS-----KSS--ARGQ----- 1207  
Db 2178 IANGYPSGTNSFGAVVAYVSHKGFYIKGEKSTCEATGQSKSPRTCHPVSQNEPRAVE 2237  
Qy 1208 -AFLKHLTDLISR-----GTLFV-----PONGSLP---G 1232  
Db 2238 NGFLEHTTGRTFSEARFQCNPGYKAAGSPVFCQANRHHSDAPLSCTPLNCGKPPPIQ 2297  
Qy 1233 NKSLSGRDIE-----HHLTNNVSYND-----LVN 1258  
Db 2298 NGFLKSEFEVGSVQVFCVNGYELVGNISWTQKSGKWSKPSFKCVPTKCAEPPLEEN 2357  
Qy 1259 GTFPLRTMIGSOLLTFPSODOLHOETREFVDGRSIL-----QMD----- 1295  
Db 2358 QLVKELASEVGMTISKKEGHA---LQGPSVLKCLPBGQWNGSFPICKMVLGCSPLI 2413  
Qy 1296 ---IIANGLIHISEP-----LRAPPTAATAHSGIGTGI-FCAYV-----LV 1335

Db 2414 PFGVASSGALHFGSTYKYLICVDGFEFLKSGPTLICOADSTWSSPLPECVPEVCEQPEIL 2473  
Qy 1336 TGAIALAASYFRLLKORT--TGFORFDOKRILM-----SMLASSSPRISOTLCAHPQ 1386  
Db 2474 NGIITHVGLATLSTTITCKRKGFELVGNATILCGENQWL--GSKPMKPIECEBP 2528

RESULT 14  
PCT-US02-09671-698  
Sequence 698, Application PC/TUS0209671  
GENERAL INFORMATION:  
APPLICANT: Zycoos Inc.  
TITLE OF INVENTION: TRANSLATIONAL PROFILING  
FILE REFERENCE: 08191-026W01  
CURRENT APPLICATION NUMBER: PCT/US02/09671  
CURRENT FILING DATE: 2002-03-28  
PRIOR APPLICATION NUMBER: 60/279,495  
PRIOR FILING DATE: 2001-03-28  
PRIOR APPLICATION NUMBER: 60/292,544  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: 60/310,801  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: 60/326,370  
PRIOR FILING DATE: 2001-10-01  
PRIOR APPLICATION NUMBER: 60/336,780  
PRIOR FILING DATE: 2001-12-04  
PRIOR APPLICATION NUMBER: 60/358,985  
PRIOR FILING DATE: 2002-02-20  
NUMBER OF SEQ ID NOS: 2041  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 698  
LENGTH: 3623  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US02-09671-698

Query Match 4.3%; Score 338; DB 1; Length 3623;  
Best Local Similarity 31.4%; Pred. No. 9.9e-11;  
Matches 101; Conservative 30; Mismatches 125; Indels 66; Gaps 21;  
Qy 182 SVFIQCQPCQCVRTITTRACWLASLAHNAKPAEGEKKMAGASVWDVNCGTGCGGIG 241  
Db 153 SFCICPPQ-----WKGPL-----CSADVNECEIYSGTFLSCQNG-GTCVNTMG 195  
Qy 242 FNGTACETCTEGKYGKIH-----DQACSVHGRC-----SQGPLGDSGCDQDYGW- 286  
Db 196 ---SYSCH-CPEYTGQPCASKYKDDCEGGSVAVRCHVIGICEDLMREBAGRPKXSCVCDAGM 252  
Qy 287 ---RGVKCDMEITTDNCN---GTCHTSANCLLDPDGKASC-KCAAGFRNGTVCATAIAC 339  
Db 253 SSPNSPACTLD--RDECSFQPGPCSTLVQC-FNTQGSFYCGACPTGMOGNGYICEDINEC 309  
Qy 340 ETSNGCGST--KADCKRTTPGNRVC-VCKAGYTGIGIYCLINPCLLEHHGCDRRAEQTQ 396  
Db 310 EINNCGCSVAPPVECV-NTPESSHCQACPQYQGGKRGVCTLTIDCSVSNNGGCHPDASCSS 368  
Qy 397 TGNPNQAVNCLPKYTGDG---KYVCSLINVCLTN---MGCSFPAFCVYTBODQRICTCK 449  
Db 369 TLGSLPTCTCLPGYTGNGYGRNGCYQLSNICLSHPLCLNGQCID-TVSGY-----FKACD 421  
Qy 450 PDYTGDIYCRGSIYIGELPKNP 471  
Db 422 SGWT--GVNCTENI-NECLSNP 440  
RESULT 15  
PCT-US02-09671-701  
Sequence 701, Application PC/TUS0209671  
GENERAL INFORMATION:  
APPLICANT: Zycoos Inc.  
TITLE OF INVENTION: TRANSLATIONAL PROFILING  
FILE REFERENCE: 08191-026W01



```
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/336,760
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 701
; LENGTH: 3623
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671-701
```

```
Query Match 4.3%; Score 338; DB 1; Length 3623;
Best Local Similarity 31.4%; Pred. No. 9.9e-11;
Matches 101; Conservative 30; Mismatches 125; Indels 66; Gaps 21;
```

```
OY 182 SVEICGPOCVKRTITTRACWLASLHNAKPARGEVKMCALGTASVWDGVNGTGCQCGLG 241
   |||||
Db 153 SFECICPPQ-----WKGPL-----CSADVNECEIYSGTPI,SCONG--GTCVNTMG 195

OY 242 FNGTACETCTEGKYG,HC-----DQACSVHGRC-----SOGPLGDGSCDCDVGW- 286
   :| | | | |
Db 196 --SYSCH-CPPEYTPQCAKSKYDCEGGSVARCVAHICEDLMREQAGEPKYSCVCDAGMM 252

OY 287 ---RGVKKCDMETTDNCN---GTCHTSANCLLDPDGKASC-KCAAGFRGNGTVC,TA,INAC 339
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 253 SSPNSPACTLD--RDECSFQPGPCSTLVQC-FNTQGSFYCGACPTGQWQNGYICEDINEC 309

OY 340 ETSNNGCGST--KADCKRTTPG,GNRVG-VCKAGYTG,GI,VCLEINPCLNENHGGCDRNAECTQ 396
   | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : |
Db 310 EINNCGGSVAPPEVCEV-NTPGSSHQACPPGYGGDGRVCTLTIDICSVSNMGCHPDASCSS 368

OY 397 TGPNOAVNCNCLPKYTG,SG---KYCSLINVC,LTN---NGGCS,PFACNYTEODORICTCK 449
   | : | : | | | | | : | : | : | | | : | : | : | | | : | : | : | : |
Db 369 TLGSLPLCTCLPGYTGNGYGP,NGCVQLSNICLSHPCLNQGCTD-IVSGY-----FCKCD 421

OY 450 PDYTG,DI,VC,RS,IT,GE,LPK,NP 471
   : | : | : | : | : |
Db 422 S,GMT--GVNCTENI-NECLSNP 440
```

Search completed: June 17, 2002, 12:34:07  
Job time: 313 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2002, 12:25:19 ; Search time 3351.98 Seconds

(without alignments)  
18949.000 Million cell updates/sec

Title: US-09-842-930A-1

Perfect score: 4706  
1 tcttaccagctactcac.....aagcaaaaaaaaaaaaaa 4706

Sequence: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Scoring table: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : EST:

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estnu:\*  
4: em\_estnu:\*  
5: em\_estnu:\*  
6: em\_estnu:\*  
7: em\_estnu:\*  
8: em\_estnu:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_hlc:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524.6	11.1	664	9	BB626584
2	447.8	9.5	462	10	BB626584
3	377.2	8.0	500	9	AA821602
4	357.6	7.6	758	10	BB626584
5	304.6	6.5	494	10	BB626584
6	295.8	6.3	385	9	BB842118
7	287.6	6.1	459	10	BB626584
8	253.2	5.4	459	10	BB626584
9	248.6	5.3	670	9	BB808042
10	245.4	5.2	1066	10	BB1146461
11	243.8	5.2	494	9	BB735524
12	242.2	5.1	439	9	BB847553
13	213.8	4.5	793	10	BB1832486
14	213.2	4.5	474	9	BB745695
15	204.4	4.3	373	10	BB745695
16	200	4.2	443	9	AA275305
17	198.4	4.2	460	10	BB447169

18	192.8	4.1	621	10	B1518863
19	190	4.0	809	9	A0080852
20	184.4	3.9	520	10	BG712246
21	179	3.8	378	9	A1852814
22	175.6	3.7	325	9	AA257161
23	167	3.5	928	10	B1906024
24	156.6	3.3	516	10	BE808086
25	153.2	3.3	245	10	B1183898
26	153	3.3	317	10	BF660624
27	148.8	3.2	570	10	BG383208
28	147.8	3.1	880	10	B1695595
29	141.4	3.0	405	9	AA138267
30	139	3.0	539	10	BG803514
31	136.4	2.9	460	10	R97483
32	127	2.7	425	9	A1272209
33	125.6	2.7	282	10	B1133987
34	123.2	2.6	634	10	BG964207
35	122.8	2.6	401	10	B1006218
36	122.8	2.6	441	9	AA389617
37	122.4	2.6	527	12	A2030878
38	122.2	2.6	490	12	A2487818
39	117.8	2.5	397	9	A1606926
40	117.4	2.5	647	9	A1598662
41	114.2	2.4	619	10	BG712275
42	113	2.4	570	9	BE233224
43	108.8	2.3	371	10	BM148089
44	108.8	2.3	380	12	A2487512
45	108.8	2.3	475	10	BE479875

## ALIGNMENTS

RESULT 1  
BB626584  
LOCUS BB626584 RIKEN full-length enriched, adult male drosophila mus  
DEFINITION musculus CDNA clone 9330210123 5', mRNA sequence.  
ACCESSION BB626584  
VERSION BB626584.1 GI:16464521  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 664)  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M., and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayashizaki,Y., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
Waki,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.



library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at [ratel.steng.uiowa.edu](http://ratel.steng.uiowa.edu). The subtraction has been previously described in (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_L1IB=UI-R-CAO  
TAG\_TISSUE=medulla  
TAG\_SEQ=GAACCG

BASE COUNT 115 a 104 c 125 g 118 t  
ORIGIN

Query Match 9.5%; Score 447.8; DB 10; Length 462;  
Best Local Similarity 99.6%; Pred. No. 3.9e-88;  
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 4203 ggagacagagatctgagacagacacctctgggggacactggtctgacatgagaag 4262
      |||
Db 462 GGAGACAGAGATCTGGAGACAGCACCCCTCGGGGGCAGTCGGTCTGCATGAGAAG 403
      |||
QY 4263 ccaagcaagcaaccacagtcacggtccagtgatctccagcccaagctgtctcatggaat 4322
      |||
Db 402 CCAGCAAGCAACACAGTCACGCGTTCCAGGGTATCCAGGCCCGCAGCTGCTCATGAT 343
      |||
QY 4323 cagtggtttaaagaatgacacactcataagaacagacatacctccttgatgaat 4382
      |||
Db 342 CAGTGTGTTTAAAGAAATACAACTCATTAAGCCAGCCATACCTCCTCTGGTTAAT 283
      |||
QY 4383 ctgagatgtgcacagagctaaagagacatgtctgctgatalaccctggggagacctccact 4442
      |||
Db 282 CTGGATGTGTGCCCGCGGCTAAGGAGCCATGTGGCTGGATACCTGGGGAGACCTCCACT 223
      |||
QY 4443 cctctgaacctataccggtggtctctcactccatatagtgtgtctgtctgtccctc 4502
      |||
Db 222 CCTGTGAGCTATACCGGTGTTCTCTCACTTCATATGCTGTGTCTGTCTGCGCTC 163
      |||
QY 4503 tctgtaccacaaactgtactctgtgtatctccatgaagtaagacccaagagcg 4562
      |||
Db 162 TCTTGTACCCCAAACTGTGACTGTGTGTATTCCTATTACCTTAACCAACCAAGGCG 103
      |||
QY 4563 gggctcacctctatgtctgtatctccagttaccagaagttaccgccaacatgtgac 4622
      |||
Db 102 GGGCTTACCTTTATGTCTGTATTCAGTAACCAAGTAACCTGCCACACATGTGTGC 43
      |||
QY 4623 tcaataaatgttttggaacaaataaaga 4653
      |||
Db 42 TCAATAAATGTTTGGAAACAATTAATAAAA 12
      |||

RESULT 3
AA821602 500 bp mRNA linear EST 17-FEB-1998
LOCUS vva10c03.r1 Soares.mammary.gland_NbMNG Mus musculus cDNA clone
IMAGE:1245316 5', mRNA sequence.
ACCESSION AA821602
VERSION AA821602.1 GI:2891470
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 500)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HIMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project

```

WashU-HIMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@wustl.edu](mailto:mouseest@wustl.edu)  
This clone is available royalty-free through LIND; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:659004  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 465.

FEATURES  
source  
1..500  
Location/Qualifiers

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1245316"  
/clone\_1ib="Soares.mammary.gland\_NbMNG"  
/sex="male"  
/tissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"

/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia)  
) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
primer [5',  
TGTATCCATCTGAAGTGGAGCGCGCGCGAATGGTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
constructed and normalized by Bento Soares and M.Fatima  
Bonaldi."

BASE COUNT 129 a 141 c 112 g 118 t  
ORIGIN

Query Match 8.0%; Score 377.2; DB 9; Length 500;  
Best Local Similarity 87.9%; Pred. No. 1.4e-72;  
Matches 434; Conservative 0; Mismatches 56; Indels 2; Gaps 2;

```

QY 1275 gaacatagcgagctgcagtcacatcttgctctgcactacactgagcaagatcaagat 1334
      |||
Db 9 GAATTAAGGGGGCTG-AGTCCATTGGCTTGTGCATATACAGAGAACAAGATCAAAAGC 67
      |||
QY 1335 atgtactcgaagcagactacacgggtgataggaaatcgtctgcggggcagcatcaagg 1394
      |||
Db 68 ATGTACTGTCAAGCCAGACTACACGGGTGATGGAATCGTCCCGTCGACGATCATTC 127
      |||
QY 1395 ggaagctcccaagaacctgagatcccaagtaactcttcagttgcaagagacatgctgt 1454
      |||
Db 128 GGAACCTCCAAAGAAATCCCTCAACGTCAGATATTTCTTCAGATTGCAGAGACATGCT 187
      |||
QY 1455 ccgagagctctgtagcctcgtcccttcacgctgttcgaccccttgctcagtccttcaa 1514
      |||
Db 188 CCAAGAGCTTGGTGGACCTGGGCCCTTCACCGGTTGTGGCTTCATGTGACTTCCTCA 247
      |||
QY 1515 tcatgagcccgagataaagacttggatcagcagggcctcattgtcccaagttcttcgcta 1574
      |||
Db 248 CAGTGAGTCCAAAGCTTAAAGCTGGGATTAACAGGGCCCTCATGTCCAGATTCTAGGTA 307
      |||
QY 1575 tcaagtgtgtggctgcacagcagctgtctgttgacaacctcaaaagtgaacaaagtggccac 1634
      |||
Db 308 TCACTGTGTGGCCTG-CACACAGCTACTGCTGGAGAACTTAAGGATATACAGAGTCCAC 366
      |||
QY 1635 gacctcccaagagagccagttccatctctgtcttcagagacactgtgttcataaaca 1694
      |||
Db 367 AACCTGCCAAGAGAGAGCAATTTCATCTCTGTCTCAGGACACTGCTCATTAACAA 426
      |||
QY 1695 tgaagcgaagtctcgttcagatcatcatcagcacaatggcgtcgtcatcagttataga 1754
      |||
Db 427 GAAGGGAAGGTTCTGTCCAGTGACATTATACAGACCAACGAGATCATCACGTCATATGA 486
      |||

```

[illegible]

D	b		237	GATGAGTGAATGGCACACAGTGCTGTGAGTGTGGGAGGGCTTCAGGGCACAAGCGCTGC	296
O	y		742	gaaacctgcactgagggagaagtatgttatccactcgcacaaagcatgctcttgttcacat	801
D	b		297	GAGACTGCACCCGAGGGCAAGTACGGCATCCACGTGACCAGAATGTTTGTGTGTCAT	356
O	y		802	ggagagatgttgccaaggacccttggagagacgctctccttgtactgttgacctgtgtcgca	861
D	b		357	GGGAGATTCAACCAAGGACCTTGGAAGATGCTCTGTGACTGTGATGT- GGGTGGGGA	415
O	y		862	ggaattgaagtgtgacalgyagatcaccaacaacactgcacagggagacctgtcacaceagt	921
D	b		416	GAATGTGATTTGTGTGACAAATGCACCAACCAAGAACAACTGCATTTGGGACATGCCATACGAC	475
O	y		922	gccaaactgtcctctgtgatcccagacgcgcgaagaagctctgtgcaaatgttgcgaggaattccga	981
D	b		476	GCCAACTGCCTCACCAACTCAGATGATACACACTTCATGCAAGTGTGCAGAGGATTCAC	535
O	y		982	gggaatagtgaacggtcttcta- gaccatcataatgcttltgttagaccaa- caattgagagtt	1039
D	b		536	GGAAACGGGACCATCTGCACCAATCAATCAAGCTGTGATGATCAGCCATGAGAGTGTCT	595
O	y		1040	ctcaaaaaggccgactgttaaaaaaaccccccaaggagaacccggtgtgtgtgtgtcaagaca-	1098
D	b		596	CTGCCAAGCGTGCATGTATAGGAACCA- CCAGAGAGSGCGATGTGCACGTGCAAAGCAG	654
O	y		1099	ggctataccggcgagacggcatctgtgtgcttgaatatcaaccggtlittggagaacctgtgt	1158
D	b		655	GGCTAACGCGGTGATGGCATTTGTGTGCTGCAATAATACMACCCGGTGTGGAGAACATGTGT	714
O	y		1159	ggctgtgtaca 1168	
D	b		715	GGCTGTGTACA 724	
RESULT 5					
BF601840					
L	OCUS		BF601840	494 bp mRNA linear EST 25-APR-2001	
D	EFINITION		2669242 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.		
V	ERSION		BF601840		
K	EYWORDS		BF601840.1 GI:11699063		
S	RCE		.EST.		
C	ORGANISM		cow.		
R	ERENCE		Bos taurus		
A	HORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos. 1 (bases 1 to 494)		
T	ITLE		Smith,T.P.L., Gosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Petee,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and Keel,J.W.		
J	URNAL		Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle		
M	EDLINE		Genome Res. 11 (4), 626-630 (2001)		
C	OMMENT		Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel.: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt.trimmed with phred v0.980904.e. Vector identified by cross_match with the -m1nscore 18 and -minmatch 12 options. PCR Primers FORWARD: AGGAACAGCTATGACCAT BACKWARD: GTTTCAGTCACGACG Plate: 42 row: C column: 11 Seq primer: ATTTAGGTGACACTATAG. Location/Qualifiers		









FEATURES	LOCATION/Qualifiers
Y 4037	accctggcctaaagaagaacacacttggtttccagcgttttgatccaagaagagactatga 4096
Db 59	ATTGGCGCCTTAATAACAGAGACCAATGGGATCCGGCGGTTTAGTCATAAAGATCAATGA 118
Y 4097	tgctctggcctttggccaagcagcagccccaagaatalcgcacaaccccttgatgaacctc 4156
Db 119	GGCCATGGCCTTTGACACGACGACACCTTTAAAGTAT--GCAGCGCCGGTATTAAGACGTA 176
Y 4157	agcgcggcaccgccagagtgctcctctgtgaaccccttcagaaccccttgagaagacagatct 4216
Db 177	AACGGGGGACACCAAGGAGCGCTCTGGGGCCCCCTGACAAACCTTTGGAGAACGAGAGCT 236
Y 4217	ggagaacagcagaccctcgggggacactcgytccctgcaatgagaagcagcagaacaaca 4276
Db 237	GGAGAACAGTGAACCTCTTGGGGCACTAGAGTCTGATGATGTGAAGACCAAGCCACAGCCA 296
Y 4277	cag-----tcacggttcacaggtga-ttccacagcccaagctgctc 4316
Db 297	CAGCCACGCTGAGCCCTGAGCCATCAACGGTACAGGGTACACTCCCTGCTCCAGCTGTAC 356
Y 4317	atggaatcgtgtttaaagaatgacaacactcatalaagccagccatacctaaccctctcg 4376
Db 357	TTGGATCACTTTGTTTAAAGTATGATGAATCACTAGA-----AGCCATACCTCAACCCCTCG 410
Y 4377	gttaatctcgggattgtgcgccagggcgaagagccatgtgtcctgtgataactcgtgggagact 4436
Db 411	GATATCTTGAGATTTTCTCTCATGGGTAAAGGGACATGTTTACAGGACACC--AGTATACCT 469
Y 4437	ccacctctctgagcctatcacgttggtctct--cacttccatatggtgactgtgtct 4494
Db 470	CTGCTCTCATGGGCTTACACTGTGTGCTCTATGACACCTTCCAAATGGTGCTTGGGCTGTT 529
Y 4495	ctgcctctctctgtgaaccacaacactgtgactctgtgtatctccctatgacgtaaagac 4554
Db 530	ATGCCCTATCT---ATGGGTAGCTGTGACCCCTGGTATCTCTATTAACTGTAAGCAC 585
Y 4555	caagaggcgggagcttacctctatgttctgtatcttcagtaaccacaagaatgactgcacac 4614
Db 586	CAAAAGCAAGGCTTCATCCCATATGTCTGTATTCACACTACCAAGATCACTGCGCACAC 645
Y 4615	atgtgtctcaataaatgt 4633
Db 646	GGGTGCGCTCAATAATGT 664
RESULT 10	
LOCUS	B1146461 1066 bp mRNA linear EST 05-JUL-2001
DEFINITION	602913560P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5054630 5',
ACCESSION	B1146461
VERSION	B1146461
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 1066)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L14M11A8 row: g column: 15 High quality sequence stop: 492.

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
5.2%; Score 245.4; DB 10; Length 1066;	80.2%; Pred. No. 1,7e-43;	0;	81;	15;	8;	
Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
2358 gaagtgtatcttaacaacccggttaaccttcttaaggaagaagtggaaggtctgcaga-acctgt 2416	1 GAGGTGTATCTTCAACACCCGTTACCTTTTCAGAGGAGTGTGGAAGCGTGTGTGACACCCGTGT 60	1				
2417 gcaacgtgtatcttaacaaccccgagtgctgc--catgttactcttaagcagaactgtca 2474	61 GCACCTGTGTGTCTCATGTCCCGGCTGTCTGACGTGTCTTACTTACATGATGCCAGACTGTCA 120	61				
2475 ggcctgtcccttgagagacca-gatacaacgtgttaacaacccgggcatgtgccgagatctgt 2533	121 GGCCTGCCCTTGAGAGACCCAGACACACCCGTGTACACACCGGGGCGCATGTATATGATGAGT 180	121				
2534 acaacccaatgggaagatgtgactatgtcaacacccggtcttaacgggaacagctgcagctct 2593	181 ACAACCCACACGAGGACAGTGCACAAATGTCTACTGTGTTTCACAGGGGACACCTGTGAGGCTCT 240	181				
2594 gctgagcatgggaagatgttgagctgactgtcaagcccgagctgtgtccgaaga--tgagca 2651	241 GCTTGCTCTGGAGATTTGGTCTCTGACTGTCAACCTTGCGCTGCTCTGAGCAGCTGGGCA 300	241				
2652 gtgtgatatggaggatcaacagagctccgggga-gtgcctctgtgaaacagagtgagacagcg 2710	301 GTGTGATGACGAGGAGATCAACAGGCTCCGGGCAATGTGCTGTGTGAAGCAGGAGGACAGGCC 360	301				
2711 ctctgtgtgacactcccaacagctgt-altcgagtgatgtgacacactgtctgtc-----cg 2764	361 GCTTCTGTGATGAGTCCGCCACAGTTGTGATTCACAGGTTGTCATCACTGCTGTCCCGTGCA 420	361				
2765 tgcacgccaactgtgacgagaacaacacagttgtgtgtgaac--ttgaactcgaagtgga 2822	421 CAGCCCAACATCTCTATGAGCAGAACACATATGTGTGTGATACCTTGCAATTACGAAAGTGGA 480	421				
2823 cgagg 2827	481 CGGAA 485	2823				
BASE COUNT	272 a	330 c	276 g	188 t		
ORIGIN						
1. 1066	/organism="Mus musculus"					
/strain="FVB/N"						
/db_xref="taxon:10090"						
/clone="IMAGE:5054630"						
/clone_1lb="NCI_CGAP L19"						
/lab_host="DH10B (T1 phage-resistant)"						
/note="Organ: Liver; Vector: pCMV-SpORf6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."						

, Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
,A., Takahashi,F., Takaku-Akhiba,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-Length cDNAs (Akimura,T., et al.  
2001)

JOURNAL  
COMMENT  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

#### FEATURES

##### source

Location/Qualifiers  
1..494  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="F430003K15"  
/clone\_lib="RIKEN full-length enriched, 6 days neonate  
spleen"  
/tissue\_type="spleen"  
/dev\_stage="6 days neonate"  
BASE COUNT 103 a 148 c 122 g 121 t  
ORIGIN

Query Match 5.2% Score 243.8; DB 9; Length 494;  
Best Local Similarity 77.0% Pred. No. 3.5e+43;  
Matches 386; Conservative 0; Mismatches 82; Indels 33; Gaps 6;  
OY 4155 tcaagcgccgacaccccgagctccctcgtgaccccttcacagaccgctggagaacagat 4214  
Db 1 TCAAGCGCGGACGCCCGGAGCCCTCCTGTGACCCCTTCACGACTCTTGAGAAACGGGAG 60  
OY 4215 ctgagagcagcagcactctcggggagcactcgctcctgacatgagaagccagcaaac 4274  
Db 61 CTGAGAGCAGTGAAGCCTTGTGGGACTGAGAGTCTGTATGAGAGCAGCCAGCAACAC 120  
OY 4275 cagag-----tcaggttcacaggtga-ttcacagccccaactgtc 4314  
Db 121 CACAGCCACGTGAGCCGTGAGCCATCAGGGTTACAGGGTACTCCCTGCTCCACTCTGC 180  
OY 4315 tcaatgatacagtgatttaagaagacacataaagccagcatcacctcacctc 4374  
Db 181 ACTTGATCATTTGTTTAAACTATGATTAACACAG-----AGCCATTAACCTCACCCCTC 234  
OY 4375 tggtaatactgggattgtcgccagggcctaagagaccatgttgcctgatacctggggagc 4434  
Db 235 TGGTTAGTCTGGGGTGTGCTGATGGTAAAGGGACATGTTTCCAGAGACAC-AGTATAC 293

OY 4435 ctccacctccctggagccataccgctgttctct--caattccatatgtgctgtc 4492  
Db 294 CTCCTCCTCCTCTGGGCTTACCTGTGCTCTCTCACCCTTCCAATAGTGTGGCTG 353  
OY 4493 ttctgcccctctcttcttaccacaacatgtgactcgtgtaattcttcctatgaac 4552  
Db 354 TTCTCCTCCATCT-----GTGGGTACTGTGACCCGTGGTATCTCCTATTACTGTAAAGC 409  
OY 4553 accaaagcggggtcttccactcttatgtctgatactccagtaacgaagtaacctgcac 4612  
Db 410 ACCAAAGCGAAGGCTTACCTCCATATGTTCTATTCCAGTACCCAGAAAGTACCTGCAC 469  
OY 4613 acatgtgtcctaataatgt 4633  
Db 470 ACGGCTGCCCTCAATTAATGT 490

#### RESULT 12

##### BB847553

LOCUS BB847553 439 bp mRNA linear EST 26-NOV-2001  
DEFINITION BB847553 RIKEN full-length enriched, adult male kidney Mus musculus  
cDNA clone F530009M16 5', mRNA sequence.

##### BB847553

ACCESSION BB847553.1 GI:17085928  
VERSION  
KEYWORDS  
SOURCE

##### ORGANISM

##### house mouse.

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

##### REFERENCE

##### AUTHORS

1 (bases 1 to 439)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishi  
,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
,A., Takahashi,F., Takaku-Akhiba,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-Length cDNAs (Akimura,T., et al.  
2001)

#### TITLE

##### JOURNAL COMMENT

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

#### FEATURES

##### source

Location/Qualifiers  
1..439

RESULT	13
LOCUS	BI832486
DEFINITION	BI832486 793 bp mRNA linear EST 04-OCT-2001 603082227BF1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5221410 5', mRNA sequence.
ACCESSION	BI832486
VERSION	BI832486.1 GI:15944036
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 793)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .

Tissue Procurement: Life Technologies Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: L14M11356 Row: 1 Column: 19  
High quality sequence stop: 783.

Location/Qualifiers  
1. .793

BASE COUNT	194 a	220 c	208 g	171 t
ORIGIN				

Query Match	4.58;	Score 213.8;	DB 10;	Length 793;
Best Local Similarity	84.18;	Pred. NO. 1.5e-36;		
Matches 253; Conservative	0;	Mismatches 47;	Indels 1;	Gaps 1;

3292 cctctctatgcccaagaagccaaatctaccacttgctcggccgagctgctgaagatgagg 3351 y  
3293 |||||  
193 cgtctgtatgattccctgcagagcccaattacccttgcttgctcagcagggctgtgaagaccgg 252 b  
3352 cggatctgcctaacccgactcaagtatgctctccagaagatgagtgcaaaagtgtgtgagac 3411 y  
3353 |||||  
253 cgggttgccctaccaccacagccttggccctccacagactgtgctctgtgtggttggatgata 312 b  
3412 gtaagctacggatcccaaggcccaacaagaagtgtgaattgtgagatgctctcgttaccggatg 3471 y  
3413 |||||  
313 gtagactatgtgagactgagaccacacaaagatggaatgtggagattgtttgtgattggatg 372 b  
3472 aa-agatgtgaactgacactgacatgcaaggcaggtctatgtggagatgagctctctgtcagttg 3530 y  
3473 |||||  
373 aacacatgtggaactgcacactctcagaagtgaggctatgttgagagatgctctctcatgacatgg 432 b  
3531 gaacctgtctgaagtctctcattatgctcttccctctggtcacaacaattctcgaagaagatgct 3590 y  
3532 |||||  
433 gaacctgtctgacaggtcctgagatgctctccctctactacacaaattcctcatcagcagattcact 492 b  
3591 g 3591 y  
493 g 493 y

	RESULT	14
CCUS	B745695	
DEFINITION	BB745695	474 bp mRNA linear EST 15-OCT-2001
	BB75693	RIKEN full-length enriched, adult male kidney Mus musculus
	CDNA clone F530009M16.3,	mRNA sequence.

CESSION BB745695  
VERSION BB745695.1 GI:16148634  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
(bases 1 to 474)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanaaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii  
, Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,  
Makamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Salto,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
, Takahashi,F., Takaku-Akhiba,S., Tanaka,T., Tomaru,A., Toyota  
, Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
, M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagui,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura  
, Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
, Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.

FEATURES  
source

Location/Qualifiers  
1..474  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="F53009M16"  
/clone\_id="RIKEN full-length enriched, adult male kidney"  
/sex="male"  
/tissue\_type="kidney"  
/dev\_stage="adult"  
/lab\_host="SOLR"  
/note="Site\_1: XhoI; Site\_2: SstI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGCGCGCCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5'  
GAGAGAGAGAGATCCAGAGCTCAATTAAATTAAATTAACCCCCCCCC 3'].  
cDNA was cleaved with XhoI and SstI. "

BASE COUNT 101 a 135 c 119 g 119 t

ORIGIN

Query Match 4.5%; Score 213.2; DB 9; Length 474;  
Best Local Similarity 76.5%; Pred. No. 1.9e-36;  
Matches 364; Conservative 0; Mismatches 78; Indels 34; Gaps 7;

4181 ctgtgacccttcacagacccttgagaaacagatctggaggagcagcacccttggggc 4240  
|||||  
1 CTGTGACCCCTTACCGCACTCTGGAGAACGGAGCTGGAGAACAGTACCTCTTGGGGC 60  
4241 actggtctcctgacatlgagaacccagcaagcaacacag-----tc 4281  
|||||  
61 ACTGAGGTCCTGATGTGAGAGGCCACAGACGACAGCCAGTGGAGCCCTGAGCATTG 120  
4282 acggttccaggtga-ttcgaagcccaagctgtctcattgatcagttgtttaagaatg 4340  
|||||  
121 ACGGTTACACGGTGACTCCCTGCGTCAGCTGCACTTGGATCATTTGTTTAAAGTATG 180  
4341 acaacactcataagcccaactcaactcaacctctgtgta-atctggagatgcagag 4399  
|||||  
181 ATACTCAGA-----AGCCATACCTCACCCCTTGTTAGGTCTGGGGTTGCTGCATG 234  
4400 gctaagagccatgttgcctgataacctgaggagacctccacctctctgagcctataccg 4459  
|||||  
235 GGTAGGGGACATGTTTCCAGACACC-AGTATACCTGCTGCTCTGGGCTTCACTG 293  
4460 tggctctc--cactccatatggtgtgtgtctgttctgcctctcttgaaccacaa 4517  
|||||  
294 TGGCTCTCTGCACCTTCCAGAGGTGCTGGCCTTCTGCCCCATCT---GTGGGTAG 349  
4518 ctgtgacctgtgtgattctcctctatlgacgtaagcacaagggggcttaacctta 4577  
|||||  
350 CTGTACCCCTGGTGTATCTCTCTATTACTGTAGACCAAGCAAGGCTTATCTCCATA 409  
4578 tgtctgtatlcacatgccagaagtaacctgccaacatgtgtgtcataaattg 4633  
|||||  
410 TGTCTGTATTCATTACCCAGAACTACCTGCCACAGGGGTGCTGCATTAATATG 465

RESULT 15  
T91781/c 373 bp mRNA linear EST 22-MAR-1995  
LOCUS yd52d09.s1 Soares fetal liver spleen 1NPLS Homo sapiens cDNA clone  
DEFINITION IMAGE:111857 3' similar to SP:ZK783.1 CE00760 ;, mRNA sequence.  
ACCESSION T91781 GI:723694  
VERSION T91781.1 GI:723694  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 373)  
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman  
, M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,  
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston  
, R., Williamson,A., Wohlmann,P. and Wilson,R.  
The Washu-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 710  
High quality sequence stops: 283 Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 710 Std Error: 0.00  
Seq primer: -21m13  
High quality sequence stop: 283.  
Location/Qualifiers  
1..373  
/organism="Homo sapiens"  
/db\_xref="GDB:467474"  
/db\_xref="taxon:9606"  
/clone="IMAGE:111857"  
/clone\_id="Soares fetal liver spleen 1NPLS"



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:26:34 ; Search time 31.39 seconds  
(without alignments)  
4380.498 Million cell updates/sec

Title: US-09-842-930A-2  
Perfect score: 7861  
Sequence: 1 SLPSLLTRLEQMPDYSIFRQ.....WGHGDPMSQATTVTPR 1431

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR-71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4447.5	56.6	1069	2	T42681
2	510	6.5	2871	2	A55624
3	508.5	6.5	2907	2	A57278
4	487.5	6.3	3002	2	A47221
5	496.5	6.3	1620	2	T27283
6	496.5	6.3	2918	2	A54105
7	492	6.3	2555	2	A40043
8	487	6.2	2524	2	A35844
9	484	6.2	2703	1	A24420
10	481	6.1	2871	2	A5567
11	480.5	6.1	1111	2	T26972
12	475.5	6.0	1203	2	A49175
13	474.5	6.0	2352	2	T30201
14	473.5	6.0	2437	2	SA2612
15	473	6.0	1574	2	T13954
16	473	6.0	2318	2	S45306
17	469	6.0	2321	2	S78549
18	465	5.9	2471	2	A49128
19	452	5.7	2531	2	T31070
20	446	5.7	2531	2	S18188
21	445	5.7	2531	2	A46019
22	437.5	5.6	1964	2	T09059
23	434	5.5	4135	2	T42629
24	432	5.5	3566	1	A40701
25	421	5.4	4006	2	T09070
26	413.5	5.3	1408	2	S16148
27	411.5	5.2	1722	2	B99753
28	403	5.1	1064	2	A40136
29	399	5.1	3507	2	T34513

30	392	5.0	1220	2	A56136	jagged protein pre
31	375.5	4.8	1584	2	T22674	hypothetical prote
32	373.5	4.8	1106	2	T18739	hypothetical prote
33	371	4.7	3051	2	SA2373	hypothetical prote
34	359.5	4.6	1372	2	T25933	hypothetical prote
35	357.5	4.5	1133	1	EGRT	epidermal growth f
36	355.5	4.5	3106	1	S53868	laminin alpha-2 ch
37	348.5	4.4	1786	1	MMMSB1	laminin beta-1 cha
38	348.5	4.4	2139	2	A35672	crumbs protein - f
39	348	4.4	1801	1	MMRTS	laminin beta-2 cha
40	348	4.4	4544	1	S02392	alpha-2-macroglobu
41	344	4.4	1786	1	MMHUR1	laminin beta-1 cha
42	344	4.4	3084	1	MMMSA	laminin alpha-1 ch
43	343	4.4	3075	2	S14458	laminin alpha-1 ch
44	338	4.3	3623	2	T09456	intrinsic factor-B
45	335	4.3	3672	2	T23433	hypothetical prote

## ALIGNMENTS

## RESULT 1

T42681 hypothetical protein DKFzp434E0321.1 - human (fragments)

C:Species: Homo sapiens (man)

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000

C:Accession: T42681

R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, November 1999

A:Reference number: 222233

A:Accession: T42681

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-870:871-1069 <AA>

A:Cross-references: EMBL:AL133021

A:Experimental source: adult testis; clone DKFzp434E0321

A:Note: the cDNA sequence contains a +1 frameshift near codon 870

A:Genetics:

A:Note: DKFzp434E0321.1

Query Match	56.6%	Score 4447.5	DB 2	Length 1069
Best Local Similarity	73.1%	Pred. No. 2.7e-268		
Matches 798	Conservative 100	Mismatches 146	Indels 47	Gaps 6
QY	316	GAASCKCAAGFGNGTVCATINACETSNCGGCTKADCKTTGKNRYCVKAGYTGDTGTC	375	
DB	6	GAASCKCAAGFGNGTVCATINACETSNCGGCTKADCKTTGKNRYCVKAGYTGDTGTC	65	
QY	376	LEINPCLEMHGCGDNRNAECTGTGPNQAVNCCLPKYTGDKKVCSLINVCLTNNGGCSPFAF	435	
DB	66	LEINPCLEMHGCGDNRNAECTGTGPNQAVNCCLPKYTGDKKVCSLINVCLTNNGGCSPFAF	125	
QY	436	CNYTBDORICTCKPDYTGDTGIVCRGSIVGELPKNPSTSQVFQLOEHAVERIAGCPPT	495	
DB	126	CNYTBDORICTCKPDYTGDTGIVCRGSIVGELPKNPSTSQVFQLOEHAVERIAGCPPT	185	
QY	496	VAAPLSSSFNHEPRIKDMDQGLMSQVLRHYHVGCOULLDLKVTTSATTTQGPVST	555	
DB	186	VAAPLSSSFNHEPRIKDMDQGLMSQVLRHYHVGCOULLDLKVTTSATTTQGPVST	245	
QY	556	VSDQIVFVFNNEKAVSSDIITNGVIVHIDKLLSKNLLITPKDLAGRYLOWLTVAAHH	615	
DB	246	VSDQIVFVFNNEKAVSSDIITNGVIVHIDKLLSKNLLITPKDLAGRYLOWLTVAAHH	305	
QY	616	GVTFSKSLIDSGSLVITDSITPTVFWPTDKALALPBDQDFLPMQDKDKLKSYL	675	
DB	306	GVTFSKSLIDSGSLVITDSITPTVFWPTDKALALPBDQDFLPMQDKDKLKSYL	365	
QY	676	KHYVIRDSKALASDLPKRSKMTLOGSELVSCGSGSDIGELFLWEOCKRFTIHRGLDEV	735	
DB	366	KHYVIRDSKALASDLPKRSKMTLOGSELVSCGSGSDIGELFLWEOCKRFTIHRGLDEV	425	





```

QY      868  GFGFPGDCDP-----RSCSEHGOCDE-----GITGSEECICETGWATA-----   904
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1217  GSVECSGPCGALMPDRSCTDIDCCEDNPICDGGOCTNIPGEYRCILCDGFMASEDMDK  1276
QY      905  -----SCDTPTAVFAV-----CTPA-----CSVH   923
Db      1277  TCVDYNECDLNNITCLSGCENTKSFICHCDMGISGKKGTGCCTDINECEI GAHNCGRH  1336
QY      924  ATCTENNT----CVCNLNYEGSDGITTCVTVPFCCKONNGCAKVAKCSOKGTGVSSCSKRG  978
           || | || | || | || | || | || | || | || | || | || | || | || |
Db      1337  AVCT--NMAGSFKSCSPGMIGDIKCITDLDSCSNSTGHMCSQHADCKNTMGSYRLCLCDG  1394
QY      979  YKGDTYSCLIEIDPCADGV---NGSCHEHATCRMTGPGKHKECKESHV--GDGYDCEPE  1032
Db      1395  YTGDFTTCTDDECESENILCGNGQC-----LNAPGGYRECECMGCVPSADGKACE--  1445
QY      1033  QLPDLTRCLODN----GCHCPDASCADLYFDOTTVVFHLRSPLTGQYKILTFKAKAACK-  1087
           :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      1446  --DIDECSLPNICVFHTG-----NLPGLFRCEIEIG-YEL--DSGGNGCTDV  1488
QY      1088  ----EAATLATYTNQL-----SYAQAKAYHLCASGW--LESGRVAYPTTYVASOKGAWGVIV  1138
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1489  NECLDPTTCISGNCVNTGSIYCD-----CSPDFELNPTRVACVDT-RSGNCYLNIIRRG  1542
QY      1139  DYGSRAANKSEMWDVECYRKMDVNTCTCKAG  1167
           || | || | || | || | || | || | || | || | || | || | || |
Db      1543  DNGDTACSNEIG---VGYSKASCCCSLG  1567

RESULT      3
fibrillin-2 precursor - mouse
A57278
C:Species: Mus musculus (house mouse)
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 11-Jan-2000
C:Accession: A57278
R:Zhang, H.; Hu, W.; Ramirez, F
J. Cell Biol. 129, 1165-1176, 1995
A>Title: Developmental expression of fibrillin genes suggests heterogeneity of extracellular matrix proteins
A:Reference number: A57278; MUID:95263670
A:Accession: A57278
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2907 <ZHA>
A:Cross-references: GB:I39790; NID:g762830; PIDN:AAA74808.1; PID:g762831
C:Superfamily: unassigned EGF-related proteins; EGF homology
F.I1239-1274/Domain: EGF homology <EGFL>
F.I2488-2523/Domain: EGF homology <EGF>
```

	Query Match	6.5%	Score 508.5	DB 2:	Length 2907;	
	Best Local Similarity	23.5%	Pred. No. 4.7e-23;			
	Matches 296;	Conservative 123;	Mismatches 395;	Indels 445;	Gaps 87;	
QY	190	OCVRIITFRACLAHLAHNAKPAFG---EYKCALGTA SVMGVNGTGF--CGCGLGENG	244			
	:	:::::	: : :	: : :	: : :	
Db	502	KCLPTVSSTRC-RCKMNGYF-QDANGCIDVDCTSNPCSGNVGMTPTGSYYKCHAGFOR	559			
QY	245	T-----ACEFTCEGGKYGIHEDQ-ACSCVHGRC SQGLPGDS--CDDDVGWRVKCMDEIT	297			
	:	:::::	: : :	: : :	: : :	
Db	560	TTTKQACIDIDE-----CIQNGVYLCKNRQCYN---SDGSFOCTICNAGF-----ELTT	603			
QY	298	D--NC-TNGCTHTSANCL----LDPDGKASCSKAAGF--RGNGTVCTAINACT-----SN	343			
	:	:::::	: : :	: : :	: : :	
Db	604	DGNKVCDHDECTTTNMCLINGMCINEDGSKFYCKRGFIILAPGRGRTCTDVDEQGTGIGCN	663			
QY	344	GGC-----STKADCKRTPDG-----NRVCV-----	363			
	:	:::::	: : :	: : :	: : :	
Db	664	GHCINNKGFSFDCD--PPGLAVGDVGRVCVDTHMRSTCYGEIKKGVCVRPEPQAVTKSE	720			
QY	364	-----	363			
Db	721	CCCANPDVGFGEPCPCPAKKNSAEFHGLCSSIGITVDGRDINECALDPDICANGICENL	780			

0Y	364	-----CKAGYTD--GYCLENLPLEMHGGCDNRADETOQGRQAVCNCLPKYT--G	412
Db	781	RGYKRCNCSNYEPDASGRNCIDIDECLVNRLLCD--NGLCRNT--PGSSYCTCPGCVLPT	838
0Y	413	DKQVSLTNVGLTN---NGGCSPPAFCNTEQDOORICTCP--DYGGDILVCRGSYIG--	465
Db	839	ETFCEDVNECESNPCYNGACR-----NLIQSFCESPGSKLSSTJLICIDLSKIGC	891
0Y	466	-----ELPKNPSTSOYFFOLOEHAVBRELAGPGPFTVFAPLSSSFNHEBRIDKMDQO	516
Db	892	WLNIDQNECEVNINGAT-----LKSRCATLQGA-----WQSP	924
0Y	517	GLMSQVLRHYVGGQOILLDNLKAYTTSATTLQCEPPSISYSODTYVINNEAVL-----S	571
Db	935	-----CERELD-----AACPGFARIKVGTCEDV---NECEVPGVQCPN	961
0Y	572	SDIISTNGVIAHVDKLSPKMLITTPKDALGRVLQIONITVAAHNGTKFRSKLIQDSGLS	631
Db	962	GRCVSKRSPIH---CECPBELTL--DQIGRGLD---IRMHCHLKD---EDBCIHP	1008
0Y	632	V---ITDSIHFPVTVFMPD-----KALEALPPEQODELFQMDKKKLKSYLAF-	677
Db	1009	VPGKFRMDACCACGAAMGTECEBCPKPQTKEXEYETLCRPGFA---NRGDLITGRFPC	1064
0Y	678	HYIHRSKALADLPKRSAMKTLQSGSELSVSCGG-----SDIGELFLNEOMCRPI	727
Db	1065	KDINCKAPFPMQCTGYGKRCRMTI--GSFKRCNNGFALDMEERNCTIDECRISPDLG--	1119
0Y	728	HGGLFFDVGVAYGDC-----LIM-----NPTL--GSGRC--DFTFTFDIGE	765
Db	1120	GSGLCVNTPGSGFEBCEBGEVSESGMMXKNCMDIDECERNLILCRGTQCVNTEGSGF----	Q 1175
0Y	766	CGSGCTFTRPKCPKLAKRPGVKKKCI--VNPLEFRNRNVGCON--LCTVVIQTPRC-CH-GY-F	820
Db	1176	C-----DCLPLGHEILSPREDCVDINECSLSDNL--CRNGKCVNMIGTYQCSNCGYXA	1226
0Y	821	MPDQACP-----GGRPTCPNNNGMCRDLTYTMOGILCTHTG-----NCTAC--EL	864
Db	1227	TPDRGGCTDIDECMIMMGCGTQCTN---SEGS-----ECSGEGYALMPDRSCADIDE	1275
0Y	865	CWGHFRGPDQPRSCSEHGQCDCEITGSGCLCLETGWTASCTPPAV--FAVC--TPACS	921
Db	1280	C-----ENNPDLG--DGGQC--PIREPYGCLCYDGF--MASMOMKRICIDVNECDLNPIC	1322
0Y	922	VHATCTENN---TCVCNLNLE--GDGIT--CTVWDFCKONNGGCAKVAKCSQKTOVSCSC	975
Db	1330	MEGEC--EMTKGSFICHOLOGISYAKKTTGCTDYDECEIGHANCDMHAASCLNPVGSFKSC	1388
0Y	976	KKGYGQDQSYCILEIDPCADVGVSGCHHACRMTGSGKHKECKSHIYVGDVYDCEBDLP	1033
Db	1389	REGWGNKIKCIDIDECANGTH--QCSINACVMT--PGSTRACASBEGFTDGFPCS---D	1442
0Y	1036	LDRC-----LQDNQC-----HPDASCADL---YFGDTTV-----	1062
Db	1443	VDECEENNLNCENOCLNVPATNGCEBEMFTPASRSRQDIDECSPONICVPGCNLL	1502
0Y	1063	GVHLRSPLOQYKLTEDKAKEACA--KEAATITATYNNOLASQAKYHLCSAGWLESOR	1118
Db	1503	PGMFHCICDDG--YEL--DRTGNGNCTDIDECADPINCUNGLCVNTPGRYE--CNC-----	1551
0Y	1119	VAYPTTASQCGANVNGIYDYSGRAKSSEMDVFCYRMDVNCCTKAGYVDG--FGSC	1176
Db	1552	---PDDF---QLANTGVGVD--NRVQ-----NCYLKFPDRGDSLSON	1587

RESULT 4  
A47221  
fibrillin 1 precursor - human (fragment)  
C.Species: Homo sapiens (man)  
C.Date: 02-jun-1995 #sequence-revision 25-Apr-1997 #text-change 21-Jul-2000  
C.Accession: A47221.154355..S17064.155764: S62111: A34198  
R.Cotson, G.M.: Chalberg, S.C.: Dietz, H.C.: Charbonneau, N.L.: Sakai, L.Y.:  
Genomics 17, 476-484, 1993



```
Db 1573 KACE-----DIDECSLPNICVFCTCH-----NLPGLFRCCEICG-YEL--DRSG 1613
QY 1083 EAC 1085
Db 1614 GNC 1616

RESULT 5
hypothetical protein Y64G10A.f - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27283
R:Almscough, R.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z20356
A:Accession: T27283
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1620 <NTL>
A:Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB54471.1; CESP:Y64G10A.f
A:Experimental source: clone Y64G10A
C:Genetics:
A:Gene: CESP:Y64G10A.f
A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match 6.3%; Score 496.5; DB 2: Length 1620;
Best Local Similarity 20.3%; Pred. No. 1.3e-22;
Matches 294; Conservative 126; Mismatches 403; Indels 625; Gaps 83;

QY 129 VLEICKN-----RCNDNDIIVGECGKCSQAAPCL---ETPPLAETKRCITSTIFMKGR 181
Db 72 LIRVANCASADLCIHNGT-CVPSEHNDNEQVCECPVFTGAKQYDANEC-----MANN 124
QY 182 SVETICOPQCVFTIITR--ACWLASLAHNAKPAPG-----EYKCALGTASYWD 228
Db 125 G--GCEHCVCVITIGTYCRCH-----PGFELSGDGNNTSDIDECVANGGCS 170
QY 229 G-VNGTG-----TCQCGLG 242
Db 171 RCNVSPGGRCDPDLILHADRGTGKVTSCSTDNGGCEHECENDSNGEYRCRRCRGVF 230
QY 243 -----NGTAC---ETCTEKGKYGHCQACSCVGRCSQGPLGSGSDCVGMGVACDMH- 294
Db 231 KLSNKRSCQVYDPCFDNKG--COHCTNNHGR-----AQCQCPGFH-LSYDRRS 279
QY 295 -ITTDNC---NOTCHTSANCLIDPDGKASCKACAGFR--GNGTVC-TA1INACETSNCGCS 347
Db 280 CVDIDECAKNNGCHEFCENV-----KGTFRCKCREGYOLGRDRTCEMLGCGCVANGGCG- 334
QY 348 TKADCKRTTPGRCVCAAGY----- 368
Db 335 -OHDCYDQPDGHCYCKENGYILANDOKLCHDNISTVILHARAPRLMSYETVTCVTPML 393
QY 369 -----TG-----DGIVCLEINPCLENHGGCDNACTOGRPOA--- 402
Db 394 TCHKICMLHDSGHVOCFCDDGDELIDSKFCODINECHENNG-----DCSQCVMLAGSV 447
QY 403 VCNCLP--KYTGDKVKSILINVCILNNGGSPFAFCNUTEDORICTCKPDY--TGDCIV 458
Db 448 ECQCGPGRFLMKDKRTCDIDISCSNNAGGE--QITCS-NQEGGYNCSEPEPGLSEDSHS 504
QY 459 CRGSITGELPKNPSTSQYFLOIEHAYNR-LAGPGPFTVFAPLSSFNHPRIKDWDOOG 517
Db 505 C-----HDMNECLINNG----- 516
QY 518 LMSQVLRHVYVCGOOLLDNILKVTTSATTLGEPYISVSQDTPIINNEAKVSSDIIST 577
Db 517 -----GCAO-LCKNRKGRCCCFAGY-ILAHDEKSCVAASADPESLVIDEDY 563
QY 578 NGV--IHVIDKLSPKNLLITPKDAIGRVLONLITVVAANHGYTFKSKLIDSGLLSITLD 635
```

```
Db 564 SKVPGIDSIDEVIS--SIESYPADSPRL-----VFGRRRHVAKOVNQGTLSELTSS 616
QY 636 SIHTPTVYFWPPLDKALBEALPREQDPLFNQDNKKDKLSTLKHVIRDSKALASDLPRSA 695
Db 617 EVRTD-----PSEKCPNCP-----FGSITCOLSCSDCQNG 646
QY 696 WKTGGSELVRCGTGSD-IGELFLNBMCRFIHRGLLFQGVAVGICLLMNPPLGG-R 753
Db 647 KCSMRGSLSKDCDPSGTYGEC--EOTCRNGYWG---VDCAHKSCKLCDDPSTGSCR 700
QY 754 CDTFTTFDIPGEC--GSC---LFTPKPLKSK-----PKGVKKKCIYN 791
Db 701 CBD-----PEKCSDPGCDGFGYSQCNLCKMCDPRNGRCDPVPFGYCTCPDLYQSGCEK 754
QY 792 PLPF-----RRNVEGCONL--CTVVIQTPRCCHGYFMPDQO----- 825
Db 755 PCPHFTFGKNCNCFPCKCARENSECCDELITGKC-----RCKPGYUGHNCKRMCSPLFG 807
QY 826 -----ACPG-----PDTF-----CNRMCMCDLITP 847
Db 808 ACCAMKSCSPAGIRCDPVTGDCCTKKCPAGYQGNLQDQPCPAGYFGYDEQKCSKADVASP 867
QY 848 M----- 848
Db 868 HSKYCHHVTGCTCLPKGTGPLCDQCLFVETIEFDIAFSINVTACAPNTYGRPCATHTC 927
QY 849 -----GQCLCHTFNGTAC-ELCWHGRGPPDC-QPRSCSEHQGDEGITTSGE 894
Db 928 SCVNAKACDESDGSHCTPGFYGATCSFVCPGRGIDOMOLCKONGALCD--TSNGS 984
QY 895 CLCETGTAAASCDTPTAFVACTPA-----CSVAHTCF-----NNTCYNLNIEBDG 942
Db 985 CECAPGWSGKKCD-----KACAPGTFGKDCSKKCDKADGMHCDPSDECTCPGKKGHK 1038
QY 943 I--TCTVVDPCQNNKGCAKVAKCSQKGT---QVSCCKKGYKGDVSCIEIDPCAGV 996
Db 1039 CDETFDSCGF--GAGCKGICSCONGATCDSVYSCCECRPMWRK--KCDR--PCPDGR 1090
QY 997 NG-GCH-----EHAT--CR---MTGP-----GKHK----- 1015
Db 1091 FEGECNALICDCTTNDTSMYNPVAFARCDHVTEGECRCAPGAWTGRPDQCTSLGRHSEGRH 1150
QY 1016 -----CECKSHYVGDVDCE--PEQLPLDRCLQ-----DNQCGH--DA 1050
Db 1151 SCQGSNGASCDRVTFGCDPSPGFMKNCSEBCEBGLWGSNCMKILCLMHGECNENEGDC 1210
QY 1051 SCADLYFQDPTVGVFHLRSPLGOYKLTFRDAAKAEACAEATATATYNQLSYQAQAKYHICS 1110
Db 1211 ECIDGWTGPSL-----CPROF-----GRNCAQRC--NCKNAGASCDRKTGRCECL 1253
QY 1111 AGWLESGRAYVPTTYASQKCGANVGVIVDYSGRANKSEMDVEFCYRMKDVNCTCKAGYVG 1170
Db 1254 PGM--SGE-----HCEKSC--VSG--HYGAKCBET-----CECENGALC 1286
QY 1171 DGFS--CS 1176
Db 1287 DPISGHCS 1294

RESULT 6
fibrillin-2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 17-Nov-2000
C:Accession: A54105; S17063; S31101
R:Zhang, H.; Apfelroth, S.D.; Hu, W.; Davis, E.C.; Sanguinetti, C.; Bonadio, J.; Mecha
J. Cell Biol. 124, 855-863, 1994
A:Title: Structure and expression of fibrillin-2, a novel microfibrillar component pr
A:Reference number: A54105; MUID:94165150
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: mRNA
```

A:Residues: 1-2918 <ZHA>  
A:Cross-references: GB:U03272  
R:Lee, B., Godfrey, M., Vitale, E., Hori, H., Mattel, M.G., Sarfarazi, M., Tsipouras, P.  
Nature 352, 330-334, 1991  
A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two different  
A:Reference number: S17062; MUID:91304567  
A:Accession: S17063  
A:Molecule type: mRNA  
A:Residues: 752-1489,1791, 'GS', 1794-1796, 'QLT', 1922-1923, 'LD', 1926, 'P', 1928 <LEE>  
A:Cross-references: EMBL:X62009  
R:Miliewicz, D.M.  
Submitted to the EMBL Data Library, December 1992  
A:Reference number: S31101  
A:Accession: S31101  
A:Molecule type: mRNA  
A:Residues: 752-1407, 'R', 1409-1489,1791, 'GS', 1794-1796, 'QLT', 1922-1923, 'LD', 1926, 'P', 1928  
A:Cross-references: EMBL:X62009  
C:Genetics:  
A:Gene: GDB:FBN2  
A:Cross-references: GDB:128122; OMIM:121050  
A:Map position: 5q23-5q31  
C:Superfamily: unassigned EGF-related proteins; EGF homology  
C:Keywords: extracellular protein  
F:1-29/Domain: signal sequence #status predicted <SIG>  
F:30-2918/Product: fibrillin-2 #status predicted <MAT>  
F:1245-1280/Domain: EGF homology <EGF1>  
F:1970-2013/Domain: EGF homology <EGF>

Query Match	6.3%	Score	496.5	DB 2	Length	218;			
Best Local Similarity	-23.1%	Pred	No. 2.6e-22;						
Matches	290;	Conservative	127;	Mismatches	400;	Indels	439;	Caps	85

```

0Y 190 OCVFRIITFACVLASLASHNAKPAG---EVKKCALSTASVMGNGFTG---CCGGGFGF 244
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 KOIPIVSSRRC-ECNMNGYK-QDANGCIDVDECTSNPCTNGDCVNPFGSYTCCGCHAGFOR 565
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 245 T---ACETCTGKGYGHICDO--ACSCVHGRCSOGPLDGS--CDXDVGMRGVAKCMEITF 297
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 TPTKQACIDIDE-----CIGMGVLCKNGRCVN--SDGSFOICNAGF-----ELFT 609
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 298 D--NC--NOTGTSANCL-----LDPDGKASCKCAAGF--RGNVTVTAINACT-----SN 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 DGRKNCVDHDECTTNNCLMGCMCINEDGSFKCIKPGFVLAPNGRYCTDVDECOTPGICMN 669
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 344 GGC-----STKADCKRTTGP-----NRYCV-----363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 670 GHCIINSEGSFRDC---PGLAVGMDGRYCVDTIHMSTCYGCIKKGVYRPPGAVTKSE 726
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 364 -----363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 727 CCANPDYGFGEPCQCPKAPKNSAEFHGLSSGVGITVDGRDINECALDPDICANGICENL 786
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 364 -----CAGTITGD--GIVCLEINPCLNHHGGCDPNNAECTQTQGPDMQVNCILPKYT--G 412
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 787 RGSYRNCNCSNGSEPDASGNCIDIDECILVNRLLCD--NLGLRMT--PGSYSTCTCPGGVYFRT 844
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 413 DGRVCSLIWVCTLN---NGGCSFPFACNTYEDODRICTCKP--DYTGADGIVCGSIYGEI 467
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 845 ETEETCEDINECSNCPVYNACR-----NNGSFNCECSPGSKLSLTGLCITDLSLKGTC 897
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 468 PNPSTSOYFQLOEHAHVELAGPGPFVAPLSSSEFNHEPRIKMDDOOGLMSOVLRYHY 527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 898 WNIQDSRCEVININGATLKS-----ECCATLGAA-----WGSF-----920
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 528 VGCQQLLDNLKATYSATLLOGEPVSIYSQDTVFINNEAKVL-----SSDIITNGVTH 562
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 931 --CERCELDLT-ACPRLARIKG-----VTCEDEV---NECEVFPGVCPMNCRCVNSKGSFH 978
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 583 VIDKLKLSPEKNLITTPDALGRVLONLTTVAANNNGYKREKSLTLDSSGLTSV---IIDSIH 638
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 979 ----CECPGGLTVL--DGTGRVCLD---IRMEQCYLKMW---EDCEIHVPDGKFRMDACC 1025
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY 639 TPVIVFWPTD-----KALEALPREQODFLFNQONKDKLXKLYLFHVIIRDSKALASD 689

Db 1026 CAVGAMNCTEEDECPKPKTEKYEFLLCPRGAQF---ANNGDVLTRPPFYKDINECKARPGM 1082

QY 690 LPRASAKMTLOGSELVSRGCTG-----SDIGELFLNEQMCRTFHRGLFLDVGVA 739

Db 1083 CTYGCARKRTI--GSFKRCNSGFALDMEERNCTDIDECRISPDLC--GSGICVNTPGSF 1137

QY 740 GIDC-----LLM-----NPTL--GGRC-DTFTTPIIDPGCGSCTFTPCPL 777

Db 1138 ECEGEGIESGFMMKMKNCMDIDGGERNPFLLRGGTCVNTBESF-----QC-----DCPL 1186

QY 778 KSKPGYKKCI-YNPILPFRNRVGGQN-LCTVVIOTPRC-CH-GY-FMPDQACP----- 828

Db 1187 GHELSPSNEDEVDINECSLSDNL--CRNGKCVNMIGTYQSCNPGYQATPRQCTDIDE 1244

QY 829 -----GCPDTPCNNRGMCRDLTYPGOCCLHTGF-----NGTAC---ELCMHGRGPRDQ 876

Db 1245 CMINNGGCDTQCTN---SEGSY---ECSCSEGYALMPDGRSCADIDEC-----ENNP 1290

QY 877 RSCSEHGCGDGGITGSGSCLCETGWTAA-----SCP--TPTAFAVCTPACSVHA 924

Db 1291 DIC-DGGCC-TINPEFYKCLCYDGFMSMDKKTCLDVAECDLNSNICFGGC----- 1340

QY 925 TCTENN---TCVCNLNAYE-GDGIT-CTVVDFFCRONNGCAKVAKCSQKQTVSCSCCKG 978

Db 1341 ---ENTKGSFICHOQLGYSVAKKGTGTCTDVECEIGAHNCMDHMASCLNIPESFKSCREG 1397

QY 979 YKGDGYSCEIIEPDADGVNNGCGCHEATCRMGPRGKHKCECKSHYVGDSVDDEPDLPLDR 1038

Db 1398 WIGNIKIKIDIDECSSNGTH--QCSINACQVNP-PGSYRCACSEGGTGDGFTCS---DVDE 1451

QY 1039 C-----LDQNCQC-----HPDASCADL--YFODTTV-----GV 1064

Db 1452 CAENINLEQNCCLNVPARYKCECMGFTPAUSDSCODDIDECSPQNICVSGTCNNLPGM 1511

QY 1065 FLHNSPLGQYKLTEDKAKEACA--KEAATATYTNQSLSYAOKAKYHLCASAGLWESGRVAY 1121

Db 1512 FHCICDDG-YEFL-DRTGNCNTDIDECADPINCVCNGLCVNTPGRE-CNC----- 1557

QY 1122 PTTVASQKCGANVYIVDYSRANKSEMDVFCYMKRVNCTCKAGYVGDG-FSCS 1176

Db 1558 PEFQNLPTG---VGCVD--NRVG-----NCTLKGPRGDSLSLN 1593

RESULT 7

A40043

notch protein homolog TAN-1 precursor - human

C:Species: Homo sapiens (man)

C:Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Aug-1999

C:Accession: A40043

R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, Cell 66, 649-661, 1991

A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosome

A:Reference number: A40043; MUID:91347367

A:Accession: A40043

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: mRNA

A:Residues: 1-2555 <ELL>

A:Cross-references: GB:M73980

C:Superfamily: unsigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

F:261-292/Domain: EGF homology <EGX1>

F:494-525/Domain: EGF homology <EGF1>

F:987-1018/Domain: EGF homology <EGX2>

F:1149-1180/Domain: EGF homology <EGF>

F:1187-1218/Domain: EGF homology <EGF3>

F:1233-1264/Domain: EGF homology <EGX3>

F:1927-1959/Domain: ankyrin repeat homology <AN1>

F:1960-1992/Domain: ankyrin repeat homology <AN2>

F:1994-2026/Domain: ankyrin repeat homology <AN3>

F:2027-2059/Domain: ankyrin repeat homology <AN4>

F:2060-2092/Domain: ankyrin repeat homology <AN5>







Db 1752 RPEGVIDITGLPVDIDE-----CR----- 1771  
QY 759 TFDPEGC--GSCI-----FTPKCPILSKPKGVAKKCIYN-PLPFRNNVEGQN--LCTV 808  
Db 1772 --ELPVCENGVCIINMWSFRCPCPV-----GFFYNKLLVCDIDDCQNGPVCQ- 1819  
QY 809 VIQPRCCCH--GYMPDCQACPGSPDT--PCNNRGMCRDLYTMMGQCLHTGTNGTACE 863  
Db 1820 --RNAECINMAGSYRCDC--PGYRFTSTGQCNDNRNECOEI----- 1856  
QY 864 LCMHGRGPGDQPCPSCEHGGCCDEGITGSGECLEGTETATASCDPLPAVAVCTPACSVH 923  
Db 1857 -----PNIQS--HGQCIDTV--GSFYCLHTGTNADOTMCLIDNECERDAGN 1902  
QY 924 ATCTENNT-----CVCN-----LNYEGDITCTVDFCKONNGCAKAVAKSOKGTQVSCS 974  
Db 1903 GTC--RNTISGFNCRNGHFLSHNND--CIDVDECATGNLNCRNQCINTVSGFCQ 1957  
QY 975 CKRGYK--GDGYSCIEI-----DP--CADGVNGGCEHATGPMGPKHAKCECKSHYVD 1025  
Db 1958 CNEGEVAPDGRFCVDINECLIDPRKCAPG-----TCQNL--DGSYRCICPPGYSIQ 2007  
QY 1026 GVDEPEPLPLDRCLQDNGOCHPDASCADLYFQDTTGVFHLRSPILGYKLPFKAKKAC 1085  
Db 2008 MDKE-----DIDEEVEPEIC--ALGTCSN-----TESFKCLCPDG--FSLSTGKRC 2053  
QY 1086 AKEATATVYQQLSYAQKAKYHLCSAGWLESRAVYPTT--YASQKCGCANVVG----- 1136  
Db 2054 -----QDLRHSYCAKF--EGKCKSPKSRHNSKECCALKGEWGDPC 2096  
QY 1137 -----IVDYS-----RANKSEMMDVFCYRMKDVN-----CTCKA 1166  
Db 2097 ELCPTEPEAFQQLCPYSGIIVGPDSDAVMDDECKEPDV--CKHGQCIINTGDSYRCCEPF 2155  
QY 1167 GYVGDGFSC 1175  
Db 2156 GYIIQGNEC 2164

RESULT 11  
T26972  
hypothetical protein Y47H9C.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 17-Mar-2000  
C:Accession: T26972  
R:Harris, B.  
Submitted to the EMBL Data Library, October 1998  
A:Reference number: Z20293  
A:Accession: T26972  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1111 <FILL>  
A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4  
A:Experimental source: clone Y47H9C  
C:Genetics:  
A:Gene: CESP:Y47H9C.4  
A:Map position: 1  
A:Introns: 50/7; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1  
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 6.1%; Score 480.5; DB 2; Length 1111;  
Best Local Similarity 22.2%; Pred. No. 7.9e-22;  
Matches 251; Conservative 98; Mismatches 341; Indels 441; Gaps 70;

QY 115 NVAIDKGV--HGLEVLEI-----QKNRCDN-----NDITIRGEGCKSQSQAPCFLETK 163  
Db 43 HVCITVKTIVDDYELKKVHTFVYVNDTEQCLNPLTGFCQTEVERQ--KASYQRL--VK 97  
QY 164 PLREPRKCIYSIYFMGKRSVFGCOPQV--RTIITRAWLASIAHNAAPAGEVAKCA- 220  
Db 98 KKKYKQCCDGYQYQTDHFCLEPCNPPCKKCKIERGKC-----ECDPGYG--KYCAS 149

QY 221 -----LGNASVWDVGNCT-----GTCCGGLGFGNGTACE--TCTBKKYGIHCDQASC 265  
Db 150 SCVSGVTGLGSKSCDDEGNANCDPELGTCTISGFGERCERCKPCKMKNPKVYKSCPC 209  
QY 266 VH--GRCSQGPLGDSGCDQDVGMR-----GVKCDMEITTTDNCNCTHTSANCL 311  
Db 210 QNGKCKNK-----BKKVCSDSGWGEFCLNKCEBKFGAEKFE-----CN--CONGATC- 257  
QY 312 LDDPGKASCK-----CAAGFRMGTV--CTAIN--ACETSN-- 344  
Db 258 DNTNGKCIKCKSGYAGALCENECVGFSGGCTQKCDCLNNOCMSSSSECKCIQMTGKHC 317  
QY 345 --GGST--KADCKR--TTPG-----NRVCCKAGYTGDIYCLIEINPLEN 384  
Db 318 DIGSGRGEFGLQCKQNTCPLEFSDSNASCDATGQCQCCSSGKGP--KDEKCKDAEQ 375  
QY 385 HGG--CDRNAECTQ-----TGPNOAVCNLPKYTGDKVCSLINVCLANNNG--CSPFAFC 436  
Db 376 YGADCSKICTCVRENTLMCAPNTGFCRCKPFGYGN--CEL--ACSKDSYSPNCEKQAMC 431  
QY 437 NYTEDQ-----RICTCKPDYTGDIYGRGSIYGEHPKPNPSTSOYFPDLOEHAVRELAP 491  
Db 432 DMNNASECNPETGSCVCKPGRGT-----KNCSE-----P 460  
QY 492 GPTVYFAPILSSFNHEPRIKMDQGLMSQYLRYHVVGCCQLLDLNLKVTTSATTLQEP 551  
Db 461 CPLDFTGP--NCAHOCQ--NORG-----VCD----- 484  
QY 552 VSIYSQDTVFITNNAKVISSDIISTNGVIHYIDKLSPKMLLTPKDALGRVLOMLTV 611  
Db 485 -----GADGCKCQCDRGWTGH 499  
QY 612 AANHGYTFESKLIQDSGLSYTDSIHPTVYFMTDALEALPEQDFLENDNKKL 671  
Db 500 RCEH-----HCPADTF----- 510  
QY 672 KSYLKFHVIKSKALASDLPRASWKTLOGSELVRCGSGSDIG--ELFLNEQMRFIHNG 730  
Db 511 -----GANCEKRCCKPKIGCDPITGECTCPAGILOG 541  
QY 731 LLEFVGV--AYGIDCLLMNPPLGRCDTFTTPIRGEQSGC--ITPKPCLK--SKPRG 783  
Db 542 ANCDIGCEPEGSYGPCKLHKCVNGCKDET-----GEC--TCQPGFFGSDCSTTCSGK- 594  
QY 784 YKKKCIYNLPPRRRVEGQNIC--TVIQTGRCHGYMPQCCAPGSP-----DT 833  
Db 595 -----YG-----ESCELSQPCSDASCSKQTKC-----LCPLGKGVSCDQ 630  
QY 834 PC--NNRG--MCRDLYTPM-----GOCL--CHTFGNGTACE--LCMHGRFGPDQ--RS 878  
Db 631 KCDPNTFGLCOEIVTPSPFCASTDRKNVGLSCPPGSSGITHENHCPRAGSYDGGQYVCS 690  
QY 879 CSEHQCDEGITGSGECLEGTETATASCD--TPTAVF-----AVCPACSVNATCTE--NNT 931  
Db 691 CADGHGCD--PTTGECICEPEYHGKTCSEKCPDGKYGYGALDPCPKCASGSTCHINL 747  
QY 932 CVCNLTNEBEDGITTYVDPCKONNGGCAKAVAKSOKGTQV-----CSCKKRYKKGDSYC 986  
Db 748 CICIPAGLE--GALCTPRCSAGFWGNGCRQVCTSTYKQCNQTEGSCCPAGFQD--RC 803  
QY 987 IELDPADGVNG--GCEHNTACRMTPG-----KHKCECKSHYVGD-----GVDC 1029  
Db 804 DK--PCEDGYTPDCTLKCKKCGATSSCNRYSGACHHPRGTGEGCHALCESTFGLMC 861  
QY 1030 EPEQIPLDRCLQDNGOCHPDASCADLYFQDTTGVGFHL--RSLPGYKLTFTD 1079  
Db 862 SKF-----C--PKDGGDGYEEDDAIIGCHVDMSCGKAKQKFE 898

RESULT 12  
A49175  
Motch B protein - mouse (fragment)  
N:Alternate names: Notch homolog



C:Species: Mus musculus (house mouse)  
 C:Date: 21-Jan-1994 #sequence\_revision 05-Jan-1996 #text\_change 20-Sep-1999  
 C:Accession: A49175; PH570; S32113  
 R:Lardelli, M.; Lendahl, U  
 Exp. Cell Res. 204, 364-372, 1993  
 A:Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues  
 A:Reference number: A49175; MUID:93178563  
 A:Accession: A49175  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1203 <LAR>  
 A:Cross-references: EMBL:X68279; NID:g287989; PIDN:CAA48340.1; PID:g287990  
 A:Experimental source: embryo  
 A:Note: sequence extracted from NCBI backbone (NCBI:P126158)  
 C:Comment: This protein has many EGF repeats and 11n-12/Notch repeats.  
 C:Comment: This protein is one of the neurogenic proteins controlling the decision between C and S.  
 F:143-174/Domain: EGF homology <EGF1>  
 F:482-513/Domain: EGF homology <EGF1>  
 F:560-591/Domain: EGF homology <EGF>  
 F:674-705/Domain: EGF homology <EGX2>  
 F:712-743/Domain: EGF homology <EGF3>  
 F:836-867/Domain: EGF homology <EGX3>

Query Match 6.0%; Score 475.5; DB 2; Length 1203;  
 Best Local Similarity 23.0%; Pred. No. 1.8e-21;  
 Matches 255; Conservative 112; Mismatches 412; Indels 341; Gaps 70;

QY 219 CALGTASWMDGVNCTGTCCGGLGNGTACETCTEKGKGIHCDACS--CVH--GRCSOGPL 275  
 DB 67 CHKGLCDTNPGLNGLYICTCPQYKAGADCTEDVD-----ECAMANSNCEHAGKCVN--- 118  
 QY 276 GDGS--GDCVGMGKVCMDIEITDNCNG--TCHTSANCLDPPDKASKCAAGFRGNTV 332  
 DB 119 TDGAFHECEKLGKYGAPRCMDI--NECHSDPCQDAPTC-LDKIGFTCLCMPFRK--GVH 173  
 QY 333 C-TINNCET-----SNGGCTKADCKRTTPGNRY-CVCKAGATGDDGVC-LEINPCLLENH 385  
 DB 174 CELEVNCEQSNPCVNNQOCVDK-----NRFQCLCPFGTGP--VCQIDIDDC--SS 221  
 QY 386 GCGDRNAECTGTGTGPNQAVNCLPRTYGDGKYSLINVCLTNNGGCSPPAFGNTYE---Q 441  
 DB 222 TPCILNGAKCID-HPNGYECQCATGFTG-----ILCDEINDCDDPD--CHHQCQCGDI 271  
 QY 442 DQRTCTCKPRTYGTGDIYCRGSIYELPKRNPSTQYFFOLOHNAVRELAPGPPFYFAPLS 501  
 DB 272 DSYTCINPGFM--GAICSDOI-DECYSSPC-----LNDGRCIDLVN--GYQCNCQPGT 320  
 QY 502 SSFHEPRIKMDQOGLMSOVL-----RYHVGCOQLLDLNLKVTTSATTLQGPVSISV 556  
 DB 321 SGLNCEINFDCCASNPCMHGVCVDGINRYSCV-----CSPGFTGCRNCIDI 366  
 QY 557 SODTVFINNEAKVLSDIISTNGVIVHIDKLSPKRLITPKDALGRVQLNLTVAANHG 616  
 DB 367 DE---CASNPCKGATGINDVNGFRICPEGRHPHSCSYQNECLSNPCICNGCTGGLSG 423  
 QY 617 YTKRSKLIQDSGLSVITDSIHTPYTVFWPTDKALEALPPQOQFLFNQDNKDKLKSTLK 676  
 DB 424 Y---KCLCDAGWGVNCE-----VDKNECLSNPCQNGTGN---NLVNGY-- 462  
 QY 677 FHVIRDSKALASDLPRASMKVTLQGSLSVRCGSDIGELFLNQMRFIHRLGLPEVG 736  
 DB 463 -----KCTCKKGFKGYMCOV-----NIDECASN--C--LNQGTCPDV 497  
 QY 737 VAVGIDCLMNPILGKRCDFTFTRDIPGEG-----SCITFP-----KCP 777  
 DB 498 SGYTGCHML--PYTGKNCQVFLAPCSPNCPENAAVCKEAPNFESFCLCAFGMGQKRTV 555  
 QY 778 K-----SKP---KGV-----KKCIYNPLPFRNRVCGCN-- 804  
 DB 556 DVDEICSPCMNNGVCHNTGGSYVCECPGFGSGMDCEEDINDCLANP-----CONG 606

QY 805 -LCTVVIQTPRC--CHGYEMPD-CQA-----CPGGDFPCCNNRGMCRDLTPMGQCLHTGF 857  
 DB 607 GSCVDHVNFTFSCQCHPFGIDGKQCTDMNEC---LSEPCKNNGCTSD--YVNSYTCCTPAG 662  
 QY 858 NGTACE-----LCWGRFGP-----DCOPRSCSEHQ 884  
 DB 663 HGVCENNNIDECTESCFNGCTGVGINSFSLCPVGFTGFPFLHDINECSSNPCLNAGT 722  
 QY 885 CDEGITSGBELCTEGMTAAASCDPFAVAVCPRA--CSVAHCTENNT---CYCNLWVEG 940  
 DB 723 CVDEL--GYRCLCPGLGTTGKNQOT---LVNLCSNPKKNGCTCYEKAAPPLCLPPCW-- 776  
 QY 941 DGICTVDF--CK-----QNNGCAKVAKCSQKGTQVSCSKKYGKGDYS 985  
 DB 777 DGAVCDVLNYSKKAALQKQVPEHLQHSICINA-----GNTHHQCPGLGTYG--SYC 829  
 QY 986 CTEIDPCADGVNGGCHEHATCRMTGPKKCEKSHYVGSDVDEPRLPLDRLQDNQ 1045  
 DB 830 EEOUDECA---SNPCOHGATCN--DFIGYRCECVPGY--QGVNCEYE---VDEC--ONOP 878  
 QY 1046 CHRPASCADLYFDQTYGVFHLRSPGLQYKLTFRKAKACAKKAATATATYNQLSYAQAK 1105  
 DB 879 CONGCTCIDL-----VNHFKSCPPGTGRLCEINDECA-----CG 915  
 QY 1106 YHLCAGWLESGRVAYPTT---YASOKGANYGVYDYGSRANKSEMMDFCYRMK--DV 1160  
 DB 916 PHCLNGGCVVRIGGYTCRCLPFGAERCEBDINMCL---SNPQSE--GSLDCVQLKNY 971  
 QY 1161 NCCKKAGYV-----DGFSC 1175  
 DB 972 NCICRSAFYGRHCEFTLDVCPQKPLNGLNGTCVAVASNMMDGFTC 1014

RESULT 13  
 T30201  
 Notch homolog protein - sea squirt (Halocynthia roretzi)  
 C:Species: Halocynthia roretzi  
 C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 C:Accession: T30201  
 R:Horii, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.  
 Dev. Genes Evol. 207, 371-380, 1997  
 A:Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the A:Reference number: 220775  
 A:Accession: T30201  
 A:Status: preliminary; translated from GH/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2352 <HOR>  
 A:Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA25571.1  
 C:Genetics:  
 A:Gene: Notch

Query Match 6.0%; Score 474.5; DB 2; Length 2352;  
 Best Local Similarity 21.5%; Pred. No. 4.7e-21;  
 Matches 275; Conservative 97; Mismatches 440; Indels 467; Gaps 67;

QY 135 NRCNNPTIIVRGEG-----KCSQAPCPLEIKPLPRLFKCIYSITYFMKRSRYFICQ 188  
 DB 149 NECDTPDICOAGTCSNNDGYSCSVAGF-----EGNNECVNIDDCGHS---CQ 196  
 QY 189 --POCVRTIIRACMLASLANAKPAPGEVKMCAIGT--ASVMDGV-----NGTGT 235  
 DB 197 NGATCADAVSTYDCHC-----PAEWTGOYCTIIVDEGSLSNNAVKARRLDQTEGGFT 248  
 QY 236 CQCGKGFNGTACE-----TCTEG--KXGIHCDQAC---SC 265  
 DB 249 CNCYVAGFTRDCCSRNIDCSNVACEFHNAKCIDQAGTEFCLCTPGNRILICHDDACISDPC 308  
 QY 266 VHG--RCSGGLP--GDSGCCVDGVMGKVCMDIEITDNC-----NCTHTSANCLDP 314  
 DB 309 ARGATCDTNPLTGHMMCCPGMTDKDCSDI--DECSLGNPEEHNGCCNNT----- 359  
 QY 315 DGKASCKAAGFRGNGTVG--TAINACETSGGCTKADCKRTTPGNRYCVCKAGYTGDI 373

```
Db 360 DGESEICVAGY--SGPRCEITINECEPN--PCRNDATCLDMI--DNFNCVCMPT--GI 412
Qy 374 VC-----LEINPCLENHGGCDRNACETQGTGNQAVNCOLPRYTGDGKVCSL--INVCLTN 426
Db 413 ICDEDIDECESPPC--ANGSTCIDEV-----NAVTCSCALGETDD--CSONIDECAS 462
Qy 427 NGGCSPPAFCNVTEDODICTCKPDYTG-----DGIYCRGSIYIGELPKNPSTSQYEFQ 479
Db 463 --PCNNKATC--IDKANAAYECBCAPGYTVHCHETNIDDCVINCHGSGCRDGVNTFYCCCL 519
Qy 480 LOEHAHVRELAGBPFTVAPLSSSSNHEPRIKMDQGLMSQVLYRHVVGCOQLLDNLK 539
Db 520 LGYEGRK-----CQDTNCEASSPCENGGTCDEIIGYITCTC-----P 557
Qy 540 VTTSATLLOGEPVSISSQDVFYINNEAKYSSDIISNGVIHVHIDKLSPNNLLITPKD 599
Db 558 TGTSSSCE-----INPD 571
Qy 600 ALGRVLOMLTVAANHGYTKESKLIQDSGLSVITDSIHTEPVTPWPTDKALEALPREQ 659
Db 572 CVGNPCQYGTVCV-----DGV-----DDYSCSCTP----- 595
Qy 660 DLEFNQDKDKLSTLKHVIRDSKALASDLPRASMKTLQSGELSVRCGTG-----S 712
Db 596 -----GYTGEHCDTDINECDSPNCGMNGATCQNEVNFVCCQPPGIMGTCCSS 642
Qy 713 DIGEL-----FINEQRCFIRHGLFEDVGAVAGIDC-----LMMPTLGGRRD----- 755
Db 643 DIQESSNPCLHEYARRQOHVHCIDAGY--QGENCETELNECASPPCOHGACENKVAQV 701
Qy 756 -----FTTEDIPT-----GEGSCIFTPKCPKSKPKGKVC-----KCIYV 791
Db 702 SHCDAGYGTACEIDINECATQPCQNGTCTGSIYNCACAPKATGVNCEITELSPCVN 761
Qy 792 PLPRFRNNEGQNLCTVVIOTPRCHGTFMPDC-----QACREGPTPCNNKMGCDLYTP 847
Db 762 PC-----ENATQDESADYLAAYCQCEGFRGPFCATDINEC---VNSPCKNGGCGTNL--VP 814
Qy 848 MGOCILHTGFNGTAGCELHGHFRGPDQPRSCSEHGQCDDEGTTGCELTCEGTMTAASCD 907
Db 815 GYQCTCSGSGFGKDCDDI-----DDCSSNPCLNGGQCLDDV--GSYKCLCLDPFGGNQ 868
Qy 908 T-----PRAVPAVCT-----PACSVHATCTE--- 928
Db 869 EEVNECAFSPCKNGGICTDYVNSVCTCLSGFSLDCEKNIEDSSSCMNGGTVDGIN 928
Qy 929 NNTVCYNINTEGD-----GIT-----C 945
Db 929 SYSCSCTANFTGDKCQANVNNCASLQCONGGTCYVDSGDPKACACVHGTYGTHCESLQNIC 988
Qy 946 TVVDFCQONNGSCAKAVAKCSQKGTQVSCSKKGYKD-----GYSCIEI---DP 991
Db 989 TGPNTCK--NGS-----SCVQTSNVTSCNCLGVBGTTCANPOVSCYTGASILGTAIVSDL 1041
Qy 992 CADGVNNGCHEHATCRMTGPKHKCECKSHYVGDVDEPQPLDLRCLQDNGQCHPDAS 1051
Db 1042 CLNG--GCHDHTSTA-----HECSVAAGFTGSYCD-----IDIDECA--SVPKKNAT 1085
Qy 1052 CADLYFOFTYGVFHLBSPLOQYKILTFPKAKEDAKKATATATYQLSYAOKAKYHLCSA 1111
Db 1086 CNDL-----INSYSICALYEGATCLTLDKDECASSP-----CKN 1120
Qy 1112 GMLSEGRVAYPTTYASOKGAGNVVGIYDYSRANKSEMDVFCYR-----MKDVNCTC 1164
Db 1121 GGTCLDRL--NSFYCS--CLAGTEGVL--CEINDECEINCLNGVCLDGIIGGFSQCC 1173
Qy 1165 KAGYVGGDFSCSGNLLQYL 1183
Db 1174 PSGY--EGRRCQGDVNECL 1190
```

RESULT 14

```
s42612
transmembrane protein precursor - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Sep-1999
C:Accession: S42612
R:Bierkamp, C.; Campos-Ortega, J.A.
Mech. Dev. 43, 87-100, 1993
A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its patter
A:Reference number: S42612; MUID:94128602
A:Accession: S42612
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2437 <Bie>
A:Cross-references: EMBL:X69088; NID:9433866; PIDN:CAA48831.1; PID:9433867
C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homol
F:755-786/Domain: EGF homology <EGF1>
F:1023-1054/Domain: EGF homology <EGF>
F:1185-1216/Domain: EGF homology <EGF2>
F:1915-1947/Domain: ankryrin repeat homology <AN1>
F:1948-1980/Domain: ankryrin repeat homology <AN2>
F:1982-2014/Domain: ankryrin repeat homology <AN3>
F:2015-2047/Domain: ankryrin repeat homology <AN4>
F:2048-2080/Domain: ankryrin repeat homology <AN5>
```

Query Match 6.0%; Score 473.5; DB 2; Length 2437;  
Best Local Similarity 20.3%; Pred. No. 5,66-21;  
Matches 303; Conservative 123; Mismatches 447; Indels 617; Gaps 84;

```
Qy 135 NRCDNDITIVREGCG-----KCSQA--PCPLETKPLRETRCIYSIY 176
Db 146 NPCANG-----GQCSAFESHYICTPPNFHQTCRQDVNECAVSPSPCRMGCTINEV- 198
Qy 177 FMGRSVETIGQPO-----CVRTITTRACWLASLAHNNKRAPGEVKMALGTASVWDGVN 231
Db 199 -----GSYLCRCPEYTGPHCQR-----LYQCLPSP-----CRSGGTCYQTS-D 237
Qy 232 GTGTCCQGLGFGNGTACE-----TCTE-----GKYGIHCD-----QAC----- 263
Db 238 TTHTCSCLPFGTGYQCEHNVDDCTQHACENGSGPCIDGINTYNCHDKDKMTQGYCTEDVDE 297
Qy 264 -----SCVHGRCGSGPLGDSGDCDGVGRGVKCDMEITTDNC-----NGTCH----- 305
Db 298 CELSPNACQNGTCHNTIGGFHCVGVNMGWTGDDCSENI--DDCASAACSHATGCHDRVAS 355
Qy 306 -----TSANCLLD--DKAKSCKCAAGRGNGTCT- 334
Db 336 FFCECPHGRFTGLCHLDACISNPCQKGSNCDPTNPSGKAICTCPGY--TGSACNDID 413
Qy 335 -----AINACETSNNGCSTRA--DCK-----RTTP-----GNR 360
Db 414 ECSLGANFCEHGRCLNKTGSGFQCKLQGYEGPRCEMDVNECKSNPQONDATCLDQIGGF 473
Qy 361 VCVCAGAGTGGIYCLEIN-----PCL-----ENHGG----- 387
Db 474 HCICMPGY--EGVPC--QINSDDCASQPLNGLCKIDKINSFCECPKGFSGSLQGVYDEC 530
Qy 388 -----CDRAECTQCPNPAVNCNCLPKYTGDKVCSL--INVCLTNNGGSPAPFCYTPQD 442
Db 531 ASTPCKNGAKCTD--GPNKYTCECTPGEFGIH--CELDINECAS-----SP---CHYGCR 579
Qy 443 QRI-----CTCRPDYTGDSIVCRGSIYGLPKNPSTSQYFQLOEHAHVRELAGPFTVFA 498
Db 580 DGVASFVTCDCRPGYIGR--LEETNI--NCLSLQPCNNGTQCDREKAYICTCPKG----- 630
Qy 499 PLSSSEFNHEPRIKMDQGLMSQVLYRHVVGCOQLLDNLKVTTSATTLQGEVPSIVSQ 558
Db 631 --TTGVNCEINID--CKRRCPCDYGKCIDK----- 656
Qy 559 DTVFINNEAKYSSDIISNGVIHVHIDKLSPNNLLITPKALGRVLOMLTVAANHGYT 618
Db 657 -----INGEYCEVCEPGYSGMCNINIDCALPNCHNGGTGIDGV-----NSFTCLCPDGF- 706
```

Oy 619 KFSKILDOISGLISVITDTSIHTEPVYFWPTDKALALEALPRQOQFLFNQDDKKDKLKSVL---- 675  
 Db 707 -----RQATCLISQHNCESSNP-----CITHGSCLDQJNSTYRCVC 739  
 Oy 676 -----KFAVIDSKALASDLPRASAMKTLQSGSEL-SVGCCTGSDIGELFLNE---QMCR 725  
 Db 740 EAGWMGRNDININIELSLMPCVNGGCTCKMDTSGYLCTCAGCSGPRCMNNINIECASNPIC- 798  
 Oy 726 FTHRGLEFDVGAAYGIDICLLMNPITLGGRCDFTEFTEDIPEGC---GSC-----IFTPKC 775  
 Db 799 -INQSCSIDVAGFKNCML--PYTGEVCENVLAPCSPPRCKNGVCRESDEFOFSCNC 855  
 Oy 776 PLKSKPRKGV---KKCIYPLPRFRVBECCQL-----CTVIQTPRCHGYF----- 820  
 Db 856 PAGWQGOCTEVDINECVRNPCC--TNGVCENLIRGGFOC-----RCNPFEGALCEND 905  
 Oy 821 MEDCOACP--GG-----PDPFCNNRGMCRDL---Y 845  
 Db 906 IDDCENPNCNSGGVCODRVNGFVCVCLACGFRGERCAEDIDECVSADPCRMNGNCTDCVNST 965  
 Oy 846 TFMGOCLCTHTGNGTACEL-----CMHGRFEGPD----- 873  
 Db 966 T-----CSCPAGSGGINCELTINTPDCTESSCFNGCTCYVDGISSPSCVCLPQFTGYCQHDVN 1021  
 Oy 874 -CQPRSCSEHGOCDF-----GITG-----SGEC 895  
 Db 1022 ECDSPRCQNGSGCODDQYGTCTCPHGYTGLNQLSLVRCMDSPPCKNNGSCWQOGASFIC 1081  
 Oy 896 LCETGTAAASCTPPAVFAVCT--PACSYHATCTENNTCV-----CNLYEG----- 940  
 Db 1082 QCASGWTGTCYDVPSPVSCFEAARQOGVSAVLCRHAGOCVDAGNTHLCRQACGYTGSYCO 1141  
 Oy 941 -----DGITCTYVDF-----CKONNGCGAKVAKCSQKGTQV- 971  
 Db 1142 EQYDECOQPRPCQNGATCT--DYLGIGSCCEVPEGYHGMNKSKEINELSL-QPCONGSTCID 1198  
 Oy 972 -----SCSCKKGYKGDGYSC-IEIDPCADGV-----NGC-CHEHATCKMTGPGKH 1014  
 Db 1199 LVNTKXKSCPRGTO--GVHCEIDIDDCSPSVPLTGERPCFNGRCVDRV-----GGY 1249  
 Oy 1015 KKECKSHYVGDGVDEPBDPLPDRCLQDNGQCHPRDS--CADLYPDITTYGVFHLSPUG 1072  
 Db 1250 GCVCPAGFVGE--KEGID--VNECLSD--PCDPSGSGNYCVL-----INDPCECRGTG 1296  
 Oy 1073 ---OYKLFEDRAKEACAEATITATYNNOLSYAQAKAYHLCSAGMLESGVAVPTTYASO 1128  
 Db 1297 YTGKCEIYFNCKGKDTPCKNNGGTCAVASNTKHGYICK---COPGYGSS-----CEYDSO 1348  
 Oy 1129 KCGANYVGIYDYGSRANKSEMDVFCYRMKDYVNTCTKAGYVGDGFSC---SGNLQLOVLM 1184  
 Db 1349 SCGS-----LRCRN-CATCVSGHLSPRCLCAFGFSHGEQOTRN 1385  
 Oy 1185 SFPPLTN-----FITEVLATSSKSARQO 1207  
 Db 1386 DSPCLVNPYNGGTGTCOPISDAFFYRCSPANGLTCHILDSFSGQGR 1435  
 RESULT 15  
 T13954  
 MEGF6 protein - rat  
 C.Species: Rattus norvegicus (Norway rat)  
 C.Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
 C.Accession: T13954  
 R.Nakayama, M., Nakajima, D., Nagase, T., Nomura, N., Seki, N., Ohara, O.  
 Genomics 51, 27-34, 1998  
 A.Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs  
 A.Reference number: Z14126; MUID:98360089  
 A.Accession: T13954  
 A.Status: preliminary; translated from GB/EMBL/DBDJ  
 A.Molecule type: mRNA  
 A.Residues: 1-1574 <NAK>  
 A.Cross-References: EMBL:AB011532; NID:g3449293; PIDD:BAA32462.1; PID:g3449294  
 A.Experimental source: strain Sprague-Dawley; brain

Query Match	Best Local Similarity	22.44%	Prod. No. 3.5e-21
Matches 300;	Conservative 99;	Mismatches 460;	Indels 478; Caps
6.0% Score 473; DB 2; Length 1574;			
152 CSQAAPCLETKRPLKRETKCIY-----SIYFMGRSVFI-----	185		
DB 74 CAQQAQMC-----IGQERRTIYMSYRQVATTEARTVFRCRPGMSQKPGQEGCLSDVDEC	127		
QY 186 -----GCQPCQCVRIITRACMLASLANHAKA---PEBYMCA-LGTASWDC-----Y	230		
DB 128 ASANGACGEGPCPCNIVYGFYC-----RCRPGYOLQDGDGKTCDPVDRCRAHNGCQHRCV	180		
QY 231 NCTGT- -CGCGIGF-----NGTAC---ENCSEKXGTHGDAQ- -SCVHGRCS-----	271		
DB 181 NTPGSYLDECKRGRFLHTDGTCLALISCTVLGNGG--CQHOCVOLITVTHRCOCROYOL	238		
QY 272 -----GCPILGDS-----CCCDVGMRRVKCDMEI--TTDNCN-GTC	304		
DB 239 QEDGRKCYRSPSCAAGNCGCHITQELRGLAHCGCHGYO-LADRKTCEDYDEBALGLA	297		
QY 305 HTSANCLLDPGKASCKCAAGF--GNGTYC-----TAINACETSNCGCSTKADCKRTTP	357		
DB 298 QCAHGC-LNTQGSFKVCVCHAGYELGADQRCYRIEMELIVNSCEANGCGS--HCSHTST	354		
QY 358 GNRVCVCAGY--TGDIIVCLEINPCLENHGCCRMAECYGTG---PNAVNCMLPKY- 410			
DB 355 G-PLCTCGRYELDEDOKTCIDIDDC-----ANSPCQACANTPGVEGSCFAGYR 405			
QY 411 -TGDKVCSLINVCLTNNGG-----SPFAFCN--YTDEDORICTCKPXYTDGI 457			
DB 406 LNTDCCGEDDEDECAAGHGGCEHNCNLAGSFQCFCEAGYLDYDRGCTSLSESYVD-- 463			
QY 458 VCRGSITGELP-----KNPSTQYFQLOEH-AVRELACPGPT-VFAPLSS 502			
DB 464 -----LDGRLLFPVRLPHIAVLRLDELRLFDQDYGAEEMAAALREHETLTERFYCLDH 518			
QY 503 STHNEPRIKDMDOGLMSQVLRYHNVGCOOLLNLNKVTTSATTLQSEPPSISVSQDTVF 562			
DB 519 SFQHD-----C-SLTCCDCR--MGCTCFPGDCDCDCEPGWTGI 553			
QY 563 INNE-----AKVLSSDIITNGVTHYIDKLS-----PKNLILT-----PKDALGRVL 605			
DB 554 ICNEFCRPPDTGKNCSSPCTONG--GTCDPVLACACRPGVSAHCEHDCGPGFYGKHG 611			
QY 606 ONLTTVAANHG--YTKFSKLIQDSGLLSVITDSHTPTVTVEWPKDALEALPREQDPLF 663			
DB 612 KR-KCHCANRGRGCHLYACALCDRLGYLRCHLACRP--WAG-----PGSEELCLC 660			
QY 664 NQDN-----KDKLSYLKFHVIRDSKALASDLPRASWMT-LQGSLSVRCGTSDGE 716			
DB 661 FQSHRSCNPKD-----GSCSKRAGFGRCQAECEG----- 693			
QY 717 LFLNOMQCRFLHRIGLLPVGA-----YGIDCLLMP--TLGRCDTFT 758			
DB 694 -FFGPG-CR--HR-CTCPGVACDPVSGECRTCPGVOGEDCQCECPGTFYVNCG-- 745			
QY 759 TFDIDGEGCGSCLTFP-----KCLPKSKPKGYKKKCIYMPLRFRRNVEGQMLCTVVI 810			
DB 746 -----GSC-SCVGAARHVTGECILCPKCTYEDGADG--PEGMGILGCELTIC----- 790			
QY 811 QTPRCCH-----GYFMDQ-----AC----- 827			
DB 791 --PACHEGASCNPETGTCLCLPGFVSRCDOTCSAGWGTGQITRACANDGHCDPTGR 848			
QY 828 -----PG-----GPTD--PCN--NRGMRDLYTTPMGQCLHTGNGTAGE 863			
DB 849 CSCAAGWGLSCQARACDSGHWGPPDTHPCNNSAGHNGCDAY--SGLCLEPAGEGPRCE 905			
QY 864 -LCMHGRCGCPDQAPRSCSEH--QCDEGTTGSGECLCELTGWTAAACSD--TTPAAVAV--CTP 918			

```
Db 906 QSCROGYGPGSEQKRCCEHGAACDH---VSGACTCPAGWRGSECEHACPAAGFFGLDCDS 962
QY 919 A--CSVHATC--TENNTCVCNLNVEGD-----GITCTVYDFCKQNNNGCAKVAKC 964
Db 963 ACNCSAGAPCDNAVYGSICIPAGRWGPRCAQSCPPPLTFGLNCSQICTC--FNGASCDSTG- 1020
QY 965 SQKGTQVSCSKKGYKGDGYSCEIETPCADGVNG--GCHHATCRMTG---PGHKCECKS 1020
Db 1021 -----QCHCAPGWMGP--TCLQ--ACPPGLYKNCQHSCLCRNGGRCDPILGGCTCPE 1069
QY 1021 HYVGDGVCEPBPQLPLDKCLQDNQO---CHPDASCADLYFODTIVGVFHLRSPILGQTKL 1076
Db 1070 GMT--GLACENECIP-----GHYAAGCOLNCSLHGICDRLTG--HCLCPAG---W 1114
QY 1077 TEDKAKEACAEATATATYNNLSYAKAKYH-----LCSAGMLESG 1117
Db 1115 TGDRCQSSCVSGTFGVHCEEHACACRKGASCHHVTGACFCPPGWRGPHCEQACPRGMFGBA 1174
QY 1118 --RVAYPTTYASQ-----KCGANVVGIV-----DYGSRAN-----KSEMDVFCY 1155
Db 1175 CAQRCLCPPTNASCHHVTGECRCPPGFTGLSCBQACQPGTFGKDCHEILCQCPGETW--ACD 1232
QY 1156 RMKDVNCTCKAGYVGDG 1172
Db 1233 PASGV-CTCAAGYHGTG 1248
```

Search completed: June 17, 2002, 12:31:15  
Job time: 281 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:29:44 ; Search time 50.02 Seconds  
(without alignments)  
4949.135 Million cell updates/sec

Title: US-09-842-930A-2  
Perfect score: 7861  
Sequence: 1 SLPSLLRLRLQMPDYSIFRG.....WGHCQPMRSGQAATTVPVR 1431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organellar:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*  
15: SP-virus:\*  
16: SP-bacterioplasmid:\*  
17: SP-archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5175.5	65.8	1192	4 Q9H7H7	Q9H7H7 homo sapien
2	4447.5	56.6	1069	4 Q9UF98	Q9UF98 homo sapien
3	3762.5	47.9	897	4 Q9NRY3	Q9NRY3 homo sapien
4	2820.5	35.9	2212	4 Q93072	Q93072 homo sapien
5	2810	35.7	2570	4 Q9NRY5	Q9NRY5 homo sapien
6	603.5	7.7	2189	5 Q9B105	Q9B105 elmeria ten
7	557	7.1	3857	11 Q88840	Q88840 mus musculu
8	555.5	7.1	937	5 Q9BLJ1	Q9BLJ1 clona intes
9	536	6.8	1664	5 Q9TVO2	Q9TVO2 caenorhabdi
10	519.5	6.6	2906	11 Q9WUH9	Q9WUH9 rattus norv
11	519	6.6	2653	5 Q25253	Q25253 lucilia cup
12	499.5	6.4	2872	11 Q9WUH8	Q9WUH8 rattus norv
13	489	6.2	2470	11 Q35516	Q35516 mus musculu
14	486.5	6.2	2524	5 Q9GPA5	Q9GPA5 branchiosto
15	480.5	6.1	1111	5 Q9XWD6	Q9XWD6 caenorhabdi
16	475.5	6.0	1203	11 Q06008	Q06008 mus musculu

17	474.5	6.0	2352	5 Q61240	Q61240 halocynthia
18	473	6.0	1574	11 Q88281	Q88281 rattus norv
19	469	6.0	2281	4 Q9UPL3	Q9UPL3 homo sapien
20	469	6.0	2321	4 Q916L8	Q916L8 homo sapien
21	469	6.0	2321	4 Q9UM47	Q9UM47 homo sapien
22	465	5.9	2471	11 Q9QW30	Q9QW30 rattus sp.
23	461.5	5.9	1999	4 Q99940	Q99940 homo sapien
24	459.5	5.8	2003	4 Q00306	Q00306 homo sapien
25	453.5	5.8	3680	5 Q9VR08	Q9VR08 drosophila
26	452	5.7	2531	5 Q16004	Q16004 lytechinus
27	450.5	5.7	2471	4 Q04721	Q04721 homo sapien
28	450.5	5.7	2471	4 Q9H240	Q9H240 homo sapien
29	447	5.7	2319	11 Q9R172	Q9R172 rattus norv
30	443	5.6	2447	13 Q13149	Q13149 fuigu rubrip
31	437.5	5.6	1964	11 Q35442	Q35442 mus musculu
32	434	5.5	4135	6 Q18977	Q18977 bos taurus
33	431	5.5	3396	5 Q9VW55	Q9VW55 drosophila
34	431	5.5	4288	4 Q9NPK9	Q9NPK9 homo sapien
35	430	5.5	1140	4 Q96KG7	Q96KG7 homo sapien
36	424.5	5.4	1140	4 Q96KG7	Q96KG7 homo sapien
37	423.5	5.4	2104	5 Q964N4	Q964N4 caenorhabdi
38	421	5.4	4006	11 Q35452	Q35452 mus musculu
39	420.5	5.3	2809	4 Q96JP8	Q96JP8 homo sapien
40	415	5.3	1242	13 Q9QY57	Q9QY57 brachydanio
41	414.5	5.3	1404	5 Q9VB65	Q9VB65 drosophila
42	413.5	5.3	2104	5 Q21281	Q21281 caenorhabdi
43	411.5	5.2	1722	5 Q19350	Q19350 caenorhabdi
44	410	5.2	1193	13 Q90819	Q90819 gallus gall
45	406	5.2	1214	13 Q90YD2	Q90YD2 xenopus lae

## ALIGNMENTS

RESULT 1  
Q9H7H7 PRELIMINARY; PRT; 1192 AA.  
AC Q9H7H7;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE FLJ00112 PROTEIN (FRAGMENT).  
GN FLJ00112.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SPLEEN;  
RA Ohara O., Nagase T., Kikuno R., Okumura K.;  
RT Spleen.;  
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AK024503; BAB15793.1; -.  
DR HSSP; P98066; ITRG.  
DR InterPro; IPR000782; BIGH3\_fasciclin.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR InterPro; IPR000538; Link.  
DR Pfam; PF00008; EGF; 9.  
DR Pfam; PF02469; Fasciclin; 2.  
DR Pfam; PF00193; Link; 1.  
DR PRINTS; PR00011; EGF\_LAMININ.  
DR PRINTS; PR01265; LINKMODULE.  
DR PRODOM; PD000918; Link; 1.  
DR SMART; SM00181; EGF; 9.  
DR SMART; SM00445; LINK; 1.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 1192 AA; 128738 MW; 5966BED4B83BD9C1 CRC64;





Oy	1179	LQVLSMSPSLTNFTPEVLAFKSSSKRGCAFLLKHLDLDSIRCTLFVPQNSSLPGKKSLSG	1233
Dy	660	LQVLSMSPSLTNFTPEVLAFKSSSKRGCAFLLKHLDLDSIRCTLFVPQNSSLPGKKSLSG	719
Oy	1239	RDIHHLTNNVSFYNVDLVNGTFLRTMLGSQLITPFSODQLH-QETRFVDSRILOMDI	1297
Dy	720	RDIHHLTNNVSFYNVDLVNGTFLRTMLGSQLITPFSODQLH-QETRFVDSRILOMDI	779
Oy	1298	AANGILHIIEPLRAPPTAATAAHSGIGTGIFCAVVLYTAIALAAYSYRLKORTFO	1357
Dy	780	ASNGIIHVISRPALKAPPAPVTLHTGTGGAGIFPAILLTVGAVALAAYSYRINRRITGFQ	839
Oy	1358	RFDFCKRLMSWLAS-----SSPRISOQLCMRPQRHQSPPVPSO	1399
Dy	840	HFESEEDINVVAALGKOOPENISMPLYESTTSAPPE---PSYPFDISE	884
RESULT	4		
ID	093072	PRELIMINARY; PRT; 2212 AA.	
AC	093072;		
DT	01-FEB-1997 (TrEMBLrel_02, Created)		
DT	01-FEB-1997 (TrEMBLrel_02, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel_19, last annotation update)		
DE	KIAA0246 PROTEIN (FRAGMENT).		
GN	KIAA0246.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBTaxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BONE MARROW;		
RX	MEDLINE=97191544; PubMed=9039502;		
RA	Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.; "Prediction of the coding sequences of unidentified human genes. VI. the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain.";		
RT	DNA Res. 3:321-329(1996).		
RL	EMBL, D87433; BAIJ3377.1; -.		
DR	HSSP, P98066; ITSG.		
DR	InterPro: IPR000782; BiGRh_fasciclin.		
DR	InterPro: IPR001128; Cyt_P450.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR002049; Laminin_EGF.		
DR	InterPro: IPR000538; Link.		
DR	Pfam: PF00008; EGF_13.		
DR	Pfam: PF02469; Fasciclin; 3.		
DR	Pfam: PF00193; Xlink; 1.		
DR	PRINTS: PR01265; LINKMODULE.		
DR	PRODOM: PD000918; Link; 1.		
DR	SMART: SM00180; EGF_Lam; 1.		
DR	SMART: SM00001; EGF_Like; 13.		
DR	SMART: SM00445; Link; 1.		
DR	PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_2.		
DR	PROSITE: PS00023; EGF_1; UNKNOWN_5.		
DR	PROSITE: PS01186; EGF_2; 13.		
DR	PROSITE: PS01248; LAMININ_TYPE_EGF_2.		
DR	PROSITE: PS01241; Link; UNKNOWN_1.		
KW	EGF-like domain; Glycoprotein; Laminln EGF-like domain; Repeat.		
FT	NON_TER		
FT	NON_TER		
FT	NON_TER		
SEQUENCE	2212 AA; 237451 MW; AA95460504129134 CRC64;		

	Query Match	35.9%	Score	2820.5	DB	4	Length	2212;
	Best Local Similarity	40.2%	Pred.	1e-229;				
	Matches	563;	Conservative	219;	Mismatches	554;	Indels	65; Gaps
Oy	5	LITRLEQMPDYSIFRGYIIHYNLASAIESADAYTVFVPPNEAEIENYTRKKATSLKEDTL	64					
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :						
db	774	LLOQLDLVPAPFLPFEELLÖHHGLVQIOIEAKATYITFTFPNRBLE--AGNSNHLADLV	830					

0Y	65	RHYVLEKLLKNDJHGMREHREMLFVSLAFELLENDOLYNEAPINTVATDQVH	124
Dd	831	RHHVVLGBALSMETLKRGGHRSNLLGRPAHMYTFYHNSQPEVNHVPLRSLMEAPGRSLI	890
0Y	125	GLEYLEIQRKRCNDNDTIIVREGCKGSOQAPC---PLETKPLRETRAKTISYEMGK	180
Dd	891	GLSGVLTVGSSGRLSHSAEMALREKCVNCTRRFCTGGFLODTP---RKSCYRSGFSFS	947
0Y	181	RSVITGQPOCVRKTIITRAQWLASLHNNAKRPAEYKMKALGTASWQVNGTGTCCGL	240
Dd	948	R-----GGSYTCAKKIQVDDCCPGFFETLCEPCCGGGLGVCSGHGCCOOPRFLSGSECHHE	10030
0Y	241	GENGTACETEGEKYVGIHCDOACSCVHGRCOSQPLDDGSDCDVMMRGVCKMETMETNC	300
Dd	1004	GFIHTACEVELGKTYGNCVGTCDCAHGLIOBELGSDGSCVCNVMMOGLRDOQKRTSPQC	10633
0Y	301	NGCTHTSANCLLDPDGKASCKCAAGFRNGTVCATINACETSNGGCSTKADCKRTTPGNR	360
Dd	1064	PRKCDPARNACODSAGASTCCACAAGYSNGNIFCESEVDPAHGHGGCSPRANKTVAAPQR	11233
0Y	361	VCVCAGATGGVGICTEINCLNCLNHGSCDPMACCTGTGNQAVCNCPLRYTGDG-KVSL	419
Dd	1124	TCTQODGMGDBELCOETINCLNHGGCHHACPTTGGYQOYSCSGBESYSDGIRTCEL	11833
0Y	420	INVLJTNNGGSPFAFCNTYEDODRITCTCKPDYT-GDGIVCRGSIYELPKNPSTSOYEF	478
Dd	1184	LDPSKNNGGSPATCKSTGDDGORTCTGTAHTVGDGLTCARAYGELTIRDKHAS--FF	12414
0Y	479	QLOGHAVRELAGPPEPLVFPAR---LSSSFNHE--PRIKMDDOGLMSOYLRAHVNGCOL	533
Dd	1242	SLRLLETKELKGDOPFTTIFPRAHDLMSNLSQDLARITRANRO-----LVFKRHVYVGCRL	12366
0Y	534	LDLNLKYTTSATTLQGEPAVISVSODTVEFINNEKAVLSSDIISTNGVHVIDKLSPKNL	593
Dd	1297	RSEDLDEGAYATALSGHPLRFSEREGSITVLDNPAVSSDHNVAUNGILHFIIDRYLLPPEA	13566
0Y	594	LITRKDALGRVLOMLTVAAHNHGTYTFPSKLIQDSGLSITTSIHRTVYTPRPDKALEA	653
Dd	1357	LHMEBDAPLRPRRVNVTAAAGFGFKITFSGLLKAVAGLLPLLRASHRPFTMLMPDAAERA	14166
0Y	654	LPPEQOFLFENODKKDLKSLKLYLFENHJIRSKALASLDPRPSAKMWLQOSELSVRGCTSD	713
Dd	1417	LPRPQAMVLNHEHRODKLALIRGHMIRNVEALASDLPMLGRIMHGTPLISFC-SSTR	14755
0Y	714	IGELVPLNDOMCRFTHRGLLFDVGVAYGIDCLLMPPLGRCOTFTTFDIP-GEQSGCLTF	772
Dd	1476	PGEIEMVGGDARITQORHLPFEGGLAAGIOLLEPRLGARCDHFETRLRLNTCSIGLE	15355
0Y	773	PKCPFKSPKGVKKC-----IYNPLPRP---RNW-----EGCONCLTV	808
Dd	1536	PRCEBGSOBOGSPRACRFTPKFMTSBRPLSLGLRSVWVHPSLMGRPOGLRGCHRNCTV	15955
0Y	809	VLOTPRCCHGVMFMDCOACGBDPTPCNNKMCRLYTPRMGOSLCHTGFNGTACELMNG	868
Dd	1596	TTMKPSCCPGHGYSECQACGGRSSPCSDPGVCMDSGSCGLCSRGFAGACELCAPG	16555
0Y	869	RFGEDCQPRSCSBEHGQODEBITSSGBCLEBETGTAASCTTPAVANACTPACSVNATQTE	928
Dd	1656	AFGHPCOACRCYTHGRDEBELGGSGSCFDEBGTGRCFVOLEPRLVCTPCPAPAEVACRA	17155
0Y	929	NNTCVCMNLNAGDDITCTVVDPCFKNNGGCAKAKKSQKGTQVSCSKKGYKGDQSCIE	988
Dd	1716	GNSECELSLGTBDGRVCTVADLQODHGGCSSEHANCNSQYGTVMVCTCLPRYEGDWSQRA	17755
0Y	989	IDPCADGVNNGCHEHATCRMTGPRGKHKCECKSHYVGDVDC--EPQOLYLDRLQDNGCH	10477
Dd	1776	RNPCTDGHGGCSSEHANCISTGILNTRRCECHAGVYGDGIQCELESEBPRVDRICGPRPC	18355
0Y	1048	PDASCADYLRDDTVGVFTHLRSPRLQGYKTLTPKAKRCAKKAATLATATNOLSYAQAKYH	11077
Dd	1836	SDAMCTDLHDEKRAGVFTHLDATISGRPLGWSAEAKGACAGAVLASPRFOJLSAODQGFH	18959

Db 1836 SDANCTDLHFQEK RAGVFLHQTATSGPYGLNFSEAEAC EAGAVLASFPQLSAAQQLGFH 1895



Oy	1108	LCSAQMLESGRVAVPTTAYASQCKGANNVGYDYGSRANKSEMMDFECYRKMDYNCTCAG	1167
Dd	1896	LCIMWLANGSTAHVHVPEVPADCGMGKRGIVSLGARKNLISERWDVACRYOVACRCNG	1955
Oy	1168	YVGDDFS-CSCGNLLOVLMSFSLTFNLFLEVLAFSKSARGOAFLLHLDLSIRGLFPQP	1226
Dd	1956	FVGDEISTCNKGLLDLYLATATANFSTFYGGLLGTAYANTQRGLDFDLDELTKYKLFPV	2015
Oy	1227	NSGLPGNNLSLGRDIEHLITNVVSFYNDLVNGTFELRTMLGSOLLITSQDLHOETRFV	1286
Dd	2016	NBGFVDNMNTLSGPDELHLHSNATLLTSAN-ASQGLLPASHGISLIIDSAPDNSSWAPVA	2074
Oy	1287	DG-----RSLQMDITIAANCIIILHIISEPLRAPPTATAHS-----GLGTCIFCAVLYLVGA	1338
Dd	2075	PCTVVSVRIIYVDIAFNFIHALSPPLAPPVLAPEAPPVAAGVAILAGALLGL	2134
Oy	1339	IALAASYEFLKORTTGFOFR	1359
Dd	2135	VAGALYLARCKPMGFGESAF	2155
RESULT	5		
Q9NY15		PRELIMINARY:	PRT: 2570 AA.
ID	Q9NY15		
AC	Q9NY15:		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)		
DE	STABILIN-1.		
GN	STAB1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RA	Pollitz O., Guillot P., Grachev A., Schledzewski K., Birk R.,		
RA	Hakky N., Tebbe B., Orfanos C.E., Goerdts S.;		
RT	"Stabilin-1: an endothelial-macrophage member of the fasciclin domain		
RT	containing protein family associated with angiogenesis."		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ275213; CAB61827.1; .		
DR	HSSP; P98066; ITSG.		
DR	InterPro: IPR000782; BiGH3_fasciclin.		
DR	InterPro: IPR001128; Cyf_P450.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR002049; Laminin_EGF.		
DR	InterPro: IPR000538; Link.		
DR	pfam; PF00008; Fgf_18.		
DR	pfam; PF02469; Fasciclin; 4.		
DR	pfam; PF00193; Xlink; 1.		
DR	PRINTS; PR00011; EGFLAMININ.		
DR	PRODom; PDO00918; Link; 1.		
DR	SMART; SM00180; EGF_Lam; 1.		
DR	SMART; SM00001; EGF_like; 15.		
DR	SMART; SM00445; LINK; 1.		
DR	PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_2.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_7.		
DR	PROSITE; PS01186; EGF_2; 16.		
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 2.		
DR	PROSITE; PS01241; LINK; UNKNOWN_1.		
KW	EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.		
SQ	SEQUENCE 2570 AA; 275345 MW; 3123PABD7CEB2BP8 CRC64;		
Query Match	35.7%;	Score 2810;	DB 4; Length 2570;
Best local Similarity	40.0%;	Pred. No. 1e-228;	
Matches 561;	Conservative 222;	Mismatches 546;	Indels 72; Gaps 22;
Oy	5	LTLRLQMDYSFRCYTIHYNLASAIESDAQTVTPVPPNNEAIENITREKKATSLEDIL	64
Dd	1131	LLQQQLDVPAFSLFRELQHGLVPOIEATATITVFPTNRSL---AQGNSSHLDADTV	1187

QY	65	RHVHVUGSELKLNLDJNHNHRETMJGFSVLLAFRLRNOLYVNEARINTVATJOKVJH	124
Db	1188	RHHVVLGELALSMETLRKGGHNRSLGGLPWHVYVHNHGOPEVNHVPLBEMLEADGRSLT	1247
QY	125	GLEKVLLEOKNKNCDNNDTIIYVGECCGKCSQOAPC----PLEYKPLRETRKCIYSIYFMGK	180
Db	1248	GLISGVLTVSSSICLISHAELREKVCNCTRRRCYQGOLODTP---KRSCYYRSGFSFS	1304
QY	181	RSVFITGCOPOCVRTIITRACWLASLAHNAKRAPGEVKMICALCTASVBDVNGTJCOSGL	240
Db	1305	R-----GCSYTKAKKIQVPRDCCGPFGRFTJCEPCRGJIGVCSGHGCCODPFLSGBCJHCE	1360
QY	241	GENGTACETCTGKGVGHCDQACSCVYHGRCSGPGJLGDSCDCDVCYMRGVKCDMETTTMNC	300
Db	1361	GHEHTACAEVCEJGRYGNCTGVCDCAHGLCQDGLGDGSCVYCNWMOGJLRCDOKITTSPOC	1420
QY	301	NOTCHTNSANCLLDPROKAKSCAKCAAFRGVGTCTAIINCEYTSNGGCSYTKADCKRTTPGR	360
Db	1421	PRKOPRNACVODSSAGASTCACAGISGNGITFCSEVDPCANHGCGSPHANCTTKAAPGOR	1480
QY	361	VCVACAGYGGDITVCELEINPCLENNHGGCORNACEYOTJGNAVNCNCLPKYTTDGC-KVCSL	419
Db	1481	TCOCODGVGDELOEINISCLIHGGCHIHACEIPTGRQVYSSCSOREGSDGJRTCEL	1540
QY	420	INVCILNNGCGSPRAFONTEBODQITCTCKPRYT--GDIVYCGSIYGEJLKPXPSIYQFF	478
Db	1541	LDPCSNKNGGSPYATCKSTGCGORCTCDTJAHYVGDGLTCRAVGLLELRKHAAS--FF	1598
QY	479	OLQENAVRELABRGPTVAP---LSSSFNHE--PRIKMDQOGLMSOVLRYVHVGCOOL	533
Db	1599	SHRLLEYKELKDDPRTITVPHADLMSNLQDELAIRHNRO-----LVFRHHVVGCRLL	1653
QY	534	LIDNLTVTTSATTLQGEPRYSISVSQDPTVINEAKVLSLSDIISTNGVJHVIDIKLSLPKNL	593
Db	1654	RSEDDLEQGYATALSGHRIPLRFSBREGSIYLDNPARVYSSDHEAVNGIHFIDRVLLPREA	1713
QY	594	LITPKDALGRVLONTLTVAANHGYTKESKLIDOSGLSVITTSIHTRPVTVNPTKALEA	653
Db	1714	LHEMRDDAIPRPNVYAAAGOGFYKJFSGLLTVAGLILPRLREASHNRPMLTPDAAPRA	1773
QY	654	LBPQODELFNDONDKJLKYLPKFNHJRSKALASDLPASAMKTLQSGELSEVRCGTGD	713
Db	1774	LPRDRQWLYHEDHJKLALILGHNIRVENALASDPLMGLPLTWHGTPISFC-SKTR	1832
QY	714	IGELFLNEQMCPRINHGJLFDVGAVALIDCLMNPJLGRCSQFTTFEDP--GEQSGCIFT	772
Db	1833	PGELWGEDDARIQVNRHLRFEGGLAYGIDQLLEPELGARCDHFETRLRLMTQSIGLE	1892
QY	773	PKOPLKSKRGYKKK-----IYNLPRP---RNV-----EGCONLCTV	808
Db	1893	PRCPREGSOBQSPREACMRFYPRKFWTSPRLSLGLRSLVWVNHPSIMGRQGLGRCNHCNVT	1952
QY	809	VLOTPRSCHGTYPRQOACRGSRDTPCNKRGKROJLYTMGOSLCHTGNGTACELCHMG	868
Db	1953	TYMKPRSCGHNGHSECOACRGGSPSCSBRGVCMDSGSGOCLKSGFAGTACELCABG	2012
QY	869	RTRCPDQPRSCSGHGGCDEGTJGSGJCLEDETCWTAASCTPLPAVAFRCPACSVNATCE	928
Db	2013	ATGPRHQACRQYVHNGCDEGLGSSGSCFDEBGTGRCEVQLELQVLTCTPRAPARVCSA	2072
QY	929	NNTCVNCLNLEEGDITCTVAVDFCKONNGGCAKAVAKCSOKTOVSCSKKGYKGDYSCTE	988
Db	2073	GNSCGSESLAYEDGRVCTVAADLCQDHDGCSSEHANSOYGTWVTCILPRYDEGDGSCRA	2132
QY	989	IDPCADGVNCGCHENATCKMTGRGKJKCSCKSHYVDGYDC--EPBDLPDRCLQDNQGH	1047
Db	2133	RNPCTDGHNGGCSSEHANCSTGJNTRRCCHNGUYEDJQCLSEESPVRDRCJGORPRH	2192
QY	1048	PAWSCADLRFODPTVAVFPLRSLQGYKLTLPKAKCAGAKAATJATFYNOIYAQAKVH	1107
Db	2193	SDMCTJDOHFQERKACVYHLOATISGPTGLNFSFABEACACQAVLASPOLSAADOLGFH	2252
QY	1108	LCSAGMLESGRAVPTTYVASOKGAVVIVDYGSRANKSEMDVFCYBMKDVCNCTCKAG	1167

DB 2253 LCLMGLANGSTAHVPVPPVADCCGNGRVGVSLGARKMLSERMDAYCFRVDVACRCHNG 2312  
1168 YVGGGFS-CGSQLQVLMFSLTFLEVLAFSSKSSARGAQLKHLIDLSTRGLEFPQ 1226  
DB 2313 FVGGISITCNCKRLDLVLAATANSTFEGMLGYANATGRGLDFLDELTYKTLFV 2372  
1227 NSGLPGNSLSGRDIEHLLTFVNVSFYNDIVNGTFLRTMLGSOLLTFESODLHOETREV 1286  
DB 2373 NEGFPDNNLTGSPDELHASNATLISAN-ASGKLLPAHSGSLTIDSPAGPTNSSMAVYA 2431  
QY 1287 DG---NSILQMDIITANGILHIITSEPLRAPP-----TAATRAHSGLTGTGFCAYVLTG 1337  
DB 2432 PGTVVVSRITVWDIMAFNGIITHALASPLIAPPOQAVLAPAPVAVAGV-----AVLAG 2487  
QY 1338 AI--ALAAVSYERLKORTTGF 1356  
DB 2488 ALGLVAGALYLRANGKPTGF 2508

## RESULT 6

Q9BI05 PRELIMINARY: PRT: 2189 AA.

AC Q9BI05; 01-JUN-2001 (Tremblrel, 17, Created)  
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)  
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)  
DE MICROME PROTEIN 4.  
CS Mic4.  
OS Elmeria tenella.  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
OX Elmeria.  
NCBI\_TaxID=5802;  
RN [1]  
RC SEQUENCE FROM N.A.  
RA STRAIN=HOUTON;  
RA Tomley F.M., Billington K.J., Bumstead J.M., Clark J.D., Monaghan P.;  
RT "Emi4: A microneme protein from Elmeria tenella that contains tandem  
RT arrays of epidermal growth factor-like and thrombospondin type-1  
RT repeats.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ306453; CAC34726.1; -  
DR HSSP: P33553; IEMN.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR000884; TSP1.  
DR Pfam: PF00008; EGF\_24.  
DR SMART: SM00181; EGF\_30.  
DR SMART: SM00179; EGF\_CA\_30.  
DR PROSITE: PS00010; ASX\_HYDROXYL, 22.  
DR PROSITE: PS01186; EGF\_2; 18.  
DR PROSITE: PS01187; EGF\_CA; 18.  
DR PROSITE: PS50092; TSP1; 1.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
SEQUENCE 2189 AA; 224388 MW; C272A420B94FCB2D CRC64;

Query Match 7.7%; Score 603.5; DB 5; Length 2189;  
Best Local Similarity 24.3%; Pred. No. 2.3e-41;  
Matches 296; Conservative 114; Mismatches 459; Indels 347; Gaps 73;

QY 132 IOKNCCDNNDITIVGEGCKSQAP-----C-----PLETKPLRE--TRKCIYSIFM 178  
DB 464 MEGNACSDID-----ECSEASTEIPENCNCVNTGESSFLEKPGVELYDGKCV- KIDF- 515  
QY 179 GKRVSFICQPCQVRTITTRACMIASLAHNAKPAPEVKMCA LGTASVWDGVNGTGTCCQ 238  
DB 516 -----CAR-----GAC--NSLAHCKENPEGTALIC-----TCIA 542  
QY 239 GLGFGTACETTEGKYGIGHCDQACSCV-----HRCSCGSPGSGSCDCCVGMW-- 287  
DB 543 GYSGDGT-----QG-----HCDDIDECIAENDCTPADGGIGENTVGYSTCKCAAGYQOD 593

QY 288 GVKCDMEITITDNC--NGT--CHTSANCLIDPDGKASCCKAGFRNGITCAINACETSNG 344  
DB 594 GNSC---TIDECANGHTINCHASATC--TNGGSEFCACNAGSGVGCNDVDECTAD 649  
QY 345 GCSTKADCKRTTGCNRVCVCKAGY--TGDGIVCLEINPCLNHHGGCDRNAECTGTGPNQAV 403  
DB 650 DCGENTLLCNNTV--GSFEECTCMAFEADAATCKDKIDECASGTHCTSTACTNTA--GSFT 707  
QY 404 CNCILPKYTDGKVCGLINWCLTNNGCGSPAFPCNTBEDORI--CICKPDYGDGIVCRGS 462  
DB 708 CECNPSFDGDHKCEDVFPCCGGLHDCVNAHBCSDDNTTFKCTGIGITGEG----- 761  
QY 463 IYGLPKNPSTSQYFFLOEHAVERLAPGPPEVAPLSSFNHBPRIKMDQGLMSQV 522  
DB 762 HGE-----NCGQDIDECADQAI-----CGENTVCTNTPGSFE-----CACV 797  
QY 523 LRIYHVGCOQLLDNLKATTSATIL-----QGEVYSISVQDIYFINNEA----- 567  
DB 798 EGFVAVGAK-----LKGATSLTICIDECNDASKNTCATSADGSCKNTAGSECSCLP 851  
QY 568 -----KVLSDITSTNGVI--HVIDKLSPKNLLIPKDALGRLQNLTTVAANHGYT 618  
DB 852 GPGDGHSCTDIDECATGVCGEHA-----TCENTAGS--YNTCEA---GYT 894  
QY 619 KFSKLIQDSGLSVITDSIHTPVFWPTDK-----ALEALPPOQDFLENOQNKD 669  
DB 895 -----QODGAVGCCIDIDECASATVLPANATCVNTEGSEYFECPV---GYRHTENCT 944  
QY 670 KIKSLYKTHVLRDSKALASDLPRASAMKTLOGSELVRCGQSD--IGELFLNEOMCRFIH 728  
DB 945 K-----IDFCSEKGCNANASKENDA-----GTEAICHSQYEGNCE--GEECKMID 991  
QY 729 RELLFDVGAVGIDCLLMNPPLGRCRDTFTTEDIGEGCSCLFPKCPDLKSPK--YKK 787  
DB 992 -----EC-----SYGEPCKP-----GEGCVYDVSFGSCSCATGFIKR 1027  
QY 788 CIYNLPPRRNVGCONLCTVVIOTPRCHGFMPCOACPGPDTPCNNRGMCHDLTYP 847  
DB 1028 CT-----CQDI-----DECIDGK--NMTG--APVG-----GICNTVGS 1057  
QY 848 MQGCLCHTGF--NGTACELCHWGRGPRCQPRSCSEHQCDEGIT-----SGECLCTG 900  
DB 1058 F--TSCAAGFTGDLTCE-----DIDECATAH--TCDPNATCVNTVSGSECGCKEG 1106  
QY 901 WTAACDPTPAFAVATP---ACSVHATCTENN---TCVCNLANE--GDGITCTVVDPC 951  
DB 1107 FSGDG--HTCTIDECADPNLKNKCDPHKGIQNGTGSYTCCGCRPGVSLAADFTCDNDEC 1165  
QY 952 KONNGCAKVAKCSQKGTQVSCSCSKKGYKGDYSCTEIDPCADGVNGCCHENATCRMTGP 1011  
DB 1166 AAGTATGERSFCVDPDQSYKCEKNKGYRQCEGDVADDECEADVH--TCEHATCTNT--E 1223  
QY 1012 GKHKCECKSHYVGDVDEPEQLPDRCLQDMGCHPAPASCDLTFQDTTVVHFLRSLP 1071  
DB 1224 GSHTCNEGYOGDKCKEKTIVP---C--DNSPGNNAMC-----EATLDSYCTCKA 1272  
QY 1072 GOYKLTIPKAKPACAK--EATATATYNOLSYAOKAYH-----LCSAGWLESGRVAPPTY 1125  
DB 1273 G-----YEMKDACVYDIDECOSGTHNCDPHADCSNTDGSFTCTGSGGTYGVTLCEDE-- 1325  
QY 1136 ASQKGANVVGIVDYGRANKSEMWDVFCYRMKDYN-----CTCAGYVGDGFS 1174  
DB 1326 --DECAGNHAG-----C-----DINACTNVPSGFTCECKSGFGDGDHE 1362  
QY 1175 CSGN--LLOYLMSFBLTNLEVLAFSSKSSANG-----QAFKLKLTJLSTRGTLFVFN 1227  
DB 1363 CTEKVLILPGQIHCDSDWTAW--TECTAETKOSTRKCVALLPEVKLCPPADISAGCELGEM 1421  
QY 1228 SGLPG--NKSLSGRDIE 1242  
DB 1422 SSCPGVNNLSHRAE 1437

```
RESULT 7
088840 ID 088840 PRELIMINARY: PRT: 3857 AA.
AC 088840:
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MUTANT FIBRILLIN-1.
GN FBN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.D2;
RX MEDLINE=98069008; PubMed=9405934;
RA Bona C.A., Mural C., Casares S., Kasturi K., Nishimura H., Honjo T.,
RA Matsuda F.;
RT "Structure of the mutant fibrillin-1 gene in the tight skin (TSK)
RT mouse.";
RL DNA Res. 4:267-271(1997).
DR EMBL; AF007248; AAC62317.1; -.
DR HSSP; P35555; 1ABJ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR002557; Chitin_binding.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR InterPro: IPR000822; ZnF-C2H2.
DR Pfam; PF00008; EGF; 64.
DR Pfam; PF00683; TB; 12.
DR SMART; SM00494; ChEBD2; 2.
DR SMART; SM00179; EGF_CA; 60.
DR SMART; SM00001; EGF_like; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 61.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 50.
DR PROSITE; PS01187; EGF_CA; 61.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 3857 AA; 418301 MW; 5BC0618BC527E04C CRC64;
SQ

Query Match 7.1%; Score 557; DB 11; Length 3857;
Best Local Similarity 20.6%; Pred. No. 5e-37;
Matches 296; Conservative 136; Mismatches 403; Indels 600; Gaps 79;

QY 132 IOKNRCDNNDDTIIVGEC---GKCSQAPLPLETRKCI---YSIFMG---- 179
DB 116 IDINCELSANLCPRHRCVNLIGK--YOCACNPGYHPHTRDLFCVDIDECSTIMNGCETF 1213
QY 180 ----KRSVFICGOCPCVPTITTRACWLASLAHNAKPADEVMCALGTASVWDG-----V 230
DB 1214 CTNSDGSVECSQCPFALMPDRSC-----TDIDCE- DNPNI DGGCQCTNI 1259
QY 231 NGTGCOCGLGNGTA-CETCEGKYGIHCD-QACSVYGRCSQPLRGSGDCDVGWVG 288
DB 1260 PGEYKCLCYDGFMASEDMKTCVDVN---ECDLNPICLSGTC-ENTKSGFTCHDMGYSG 1315
QY 289 VK---C---DMEITTDNC--NGTCHTSANCLDPDGKASCSCAAGFGNGTCTAINA 338
DB 1316 KKGKGTGCDINCELCIGANCGHNAVCTMTA-----GSFKSCSPGWIGDIDKCTDDE 1368
QY 339 CETSNGGCTKADCKRTTPGNRVCYCAAGYTGDIYCLEINPLENHGGCDRNAECTQTG 398
DB 1369 CSNGTHHMSOHADCKNTM-GSYRCLCKDGYTGDFCTDIDECSENILNLCG-NGOCL-NA 1425
QY 399 PNOAVCNQ---LPKYTDDGV-----CSLINVCL-----TN 426
DB 1426 PGGYKCECDMGFVP--SADGKACEDIDECSLPNCVFGTCNHLPGIPLRCECEIGYELDRS 1483
```

```
QY 427 NGCSPFAFC-----NYTEDORICTCKPDV-----TGDGIV--CRGSIV 464
DB 1484 GGNCTDVNCELDLPPTTCISGNCVNTPGSYT-----CDCPDEFLNPTRVGCVDTCSGNTY 1537
QY 465 -----FQLOEHAVRELAGP-----GELPKNPSTSOY----- 476
DB 1538 LDIPRGNDGTPACSNELGVGVSKASCCSLGKAMQPCPELCPSVNTSEYKILCPGEGF 1597
QY 477 -----FQLOEHAVRELAGP-----GPTTVAPISSSNHPRKID-----WD 514
DB 1598 RNPPLTILIEDIDE--COELPGLCOGKCINTFGSFQCHCPGYIYNEDTRVCDVNECE 1655
QY 515 OQGLMSQVLRHVHVCQOQLL-----DNLKWTTSATLQGPVYSVSDPTVEIN 564
DB 1656 TPICGPGTCVNTGVGNYTCICPPDYMQVGNNAEYALCSSG--PGMTSAGTDI----- 1710
QY 565 NEAKVLSSDIISTNGYIHWIDKLSPKNLLTPK-----DALGRVLQNLTYVAANH 615
DB 1711 NBC-ALDPP-ICPNCIG-----ENLRGTYKCIQNSGYEVDITGKNKCVLINECVLN- 1758
QY 616 GYTKFSKLIQDSGLSVITDSITHPTV-----VFPTDKALEALPPEQDPLFND 666
DB 1759 -----SLCDNG-----QCRNTPGSEVCTCPKGFVYRPDLKTGEDIDECSSPCTNFV 1806
QY 667 NKDKLSYL-----KPHVIRDSKALASDLPBSASWKTLOGSELSVRCGTSDIGELFLNQ 722
DB 1807 CANSQSFICECSPESTLDPPTKICETIKGTCMWIV-----IDGREI--NINGATLKSE 1860
QY 723 KCRFTIHRGLLPVGVAYGIDCL--MNPTLG-----GRCDFTFTTPTDIPGEC 766
DB 1861 CCS-----SLGAAMGSPCTICQDLPICGKFSRIKGTQCEDINECEVF-----PGVC 1907
QY 767 -GSCIEFT----- 772
DB 1908 KMGCLVNSKGSFKCECPNMGTLDAATGRLCLDTRLETCFLKYDEDECTLPDIAGRHRMDACC 1967
QY 773 -----PKPLKS-----KPKGVKKCIYNPLPRFRNVGCO--NLCT- 807
DB 1968 CSVGAAMGTBECEBCECPRLNRSREYELCPRGPRFATKNDITNGAPRFPFDIDCKMIPSLCH 2027
QY 808 -----VITQTP-----RCHGY 819
DB 2028 GACRMTIGSFKCRCDSGFALDSEERNCTDIDECRISPDLCGRGQCVNTPGDECECDEGY 2087
QY 820 ---FM-----PPCQ---ACPGGPRPPCNR-----GMCERDL-- 844
DB 2088 ESEFMKMKNCMDIDECQRPDLPCRGS--ICNNTBSYRCECPRGHQLSPNISACIDINE 2144
QY 845 -----YTPMG-----QCLCHTGFGNGT-----ACE-ICWH--GR 869
DB 2145 CELSANLCPRHRCVNLIGKYOCACNPGYHPHTRDLFCVDIDECSTIMNGCETFCTNSDGS 2204
QY 870 FGPDCCP-----RSCSEHGQDE-----GITSGECLCTGTAA----- 904
DB 2205 YRCSCQPGFALMPDRSCDIDECEDNPNICDGGCCTNIPGEYRCLCYDGFMASEDMKTC 2264
QY 905 -----SCQPTTAVFAV-----CMPA-----CSVAT 925
DB 2265 VVYNECDLNPNICLSGTCTENTKGSFTICHCMDGYSGKKGCTGCTDINECTIGAHNCDRAV 2324
QY 926 CTENNT-----CVCNLNVEGDGTCTVDFCKONNGGCAKVVAKCSQKGYOVSCSKKGYK 980
DB 2325 CT--NTAGSFKSCSPGWIGDIDKCTDIDECSSGTHMGQHDCKNTKMSYGLCKIDGT 2382
QY 981 GGYSCIEIDPCADGV-----NGCCHNATCRMTGPKHKECKESHV--GDGVDCPEPOL 1034
DB 2383 GGGFTCTDIDECSENLNLGNGQC-----LNAPGGYRCECDMGFVPSADKAGE----- 2431
QY 1035 PLDRCLQDN-----GOCHPRDASCADLYFDPTVGVFHLRLSPLOKYIKTPPKAKAEC 1085
DB 2432 DIDECSLPNCVFGTCH-----NLPLRCECEIG-YEL--DRSGNC 2471
```

```

RESULT 8
ID 09BLJ1 PRELIMINARY: PRT: 937 AA.
AC 09BLJ1:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CI-METAL.
GN CI-METAL.
OS Clona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OC Clonidae; Clona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347414; PubMed=11455433;
RA Nakayama A., Satou Y., Satou N.;
RT "Isolation and characterization of genes that are expressed during
  Clona intestinalis metamorphosis.";
RL Dev. Genes Evol. 211:184-189(2001).
DR EMBL: AB041857; BAB40596.1; -.
DR HSSP: P00742; IHCG.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001212; Somatomedin_B.
DR InterPro: IPR001491; Thrombomodulin.
DR Pfam: PF00008; EGF_5.
DR PRINTS: PRO0907; THROMBOMODULIN.
DR SMART: SM00181; EGF_19.
DR SMART: SM00179; EGF_CA.
DR PROSITE: PS00010; ASX_HYDROXYL; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01187; EGF_CA_12.
DR PROSITE: PS00524; SOMATOMEDIN_B; UNKNOWN_1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 937 AA; 101043 MW; 8C67830C8E391D07 CRC64;

```

Query Match 7.1%; Score 555.5; DB 5; Length 937;  
 Best Local Similarity 20.4%; Pred. No. 7.3e-38;  
 Matches 212; Conservative 88; Mismatches 259; Indels 481; Gaps 43;

```

QY 199 ACMIASLANAPAPGEVKA-----LGTASVMDGVNGTGCOCGLGFG----- 244
DB 139 SCNNAAQPNNOQRNRYKIFCCPNPCNPSCGAYATCSNVHRYVCTCNAGTTGNPOTG 198
QY 245 -TACETCTEGKYGIHCDAQSCVHGR---CSQGPLGD----- 277
DB 199 CTAIDPCNPSGCAHA--TCSSVHRAVPCVCEAGYTGNGTCTDVNECLRPVNCPSASN 256
QY 278 -----GS--CDDCVGR--GVKCDMETTDNC--NGTCHTSAN--CLLPDGAASK 321
DB 257 NKRCVNTASSFRVCNKGKRAQSGSRC---VDINECLRPNICGINSNKRRCVNRPGG--YRCV 312
QY 322 CAAGFNGNGTCAINACTSNGGCTKADCKR--TTPGNRVGVCAAGYTGODIVCLEIN 379
DB 313 CARGYRAQSGRCVDINECLRPN--VCGPASHNKRVCNTPSFRVCNKGYRAQSGRCVDIN 371
QY 380 PCLENHGCDRNAEQTQTPNCAVNCCLRPKYTGDKVCSLIINCLINNGGCSPPAFCCNT 439
DB 372 EC--RSSPCGNNAQCIINT--PGSFTCKNGTGYTNGRACIDVNCALIRPPCSNACTNT 428
QY 440 EDDORICTCKPDYTDGIVCRGSIYELPKNPSTSOYFQLOEHAVERLAGPGPFTVFAV 499
DB 429 -PGSFTCKCKPGYTGNGLVCV-----DINECLSRACG----- 460
QY 500 LSSSFNHEPRIKMDQGLMSQVLRHYVVGCCQLLDNLKVTTSATTLQGEPIVSISVD 559
DB 461 ----- 460

```

```

QY 560 TWFINNEAKVLSIIITNGVTHVIDKLLSPKNNLITPKDALGRVLQNLITVAANHGYTK 619
DB 461 -----VNIN-----KRCINTPG----- 472
QY 620 FSKLIODSGLSVITDSITPTVTFWPTKALEALPPBOODFLFNQDNKDKLSYLFKHV 679
DB 473 ----- 472
QY 680 IRDSKALASDLPRASAMKTLQGSSELSVRCGTSIDIGELFLNEQMRFFIRGLFDVGAV 739
DB 473 -----SFR-----VCNGYR-----AQ 485
QY 740 GIDCLIMPPTLGRCDTFTTTPDI PGEGSCITTPCKPLSKRGYAKKCTIYNPLPRRV 799
DB 486 GSRVCVDIN-----EC-----RSSPCGNNAQCIINTPGSF----- 513
QY 800 EGCQNICTYVIQTPRCCHGYFMPDQACGCPDTPCCNNRGMCRDLYTPMGGLCHTFNG 859
DB 514 -----TCRCNTSY-----TGNDLICRDI-----NECEAH----- 537
QY 860 TACELCWHRFGPDPCCPRSCSEHGQCEGITSGEICLETGWTASCDTPFAVAVCTPA 919
DB 538 -----PNCGENALCINGL--GSYRICAARGTSGPLC-IDFNCAAIRP 579
QY 920 CSYHATCTE---NNTVCNMLNTEGDICTVVDPCKONNGGCAKAKVCSQKGTQVSCSK 976
DB 580 CSPNADCTFNTPGSFTCKCKPGYTGNGLVCRDINECSRPN--ACPRNQRCINTPGFNCVCA 638
QY 977 KGYKGGYSCIEIDPCADGVNNGGCHHATCRMGPCKHKCECKSHYVGVGVCPEPQDQL 1036
DB 639 IGRKRVKNCVDINECR--ASRRCIDLNASCQNT--PGSFTCTCMTGYTGNGLYC---ADI 692
QY 1037 DRCLDNGCCHPDASCADLYFDDTWTGVFHLRSPLQOYLTFEDKAKEACAKAETATYAN 1096
DB 693 NEC--NNPRACHPQATCAN-----TP--GSY-----TCN 717
QY 1097 QLSYAKKARYHLCSAGWLESGRVAYPTTYAS--OKGANVGVIVDYSRANKSEMDVFCY 1155
DB 718 -----CNRGYTGNGRVCFVNPSPSPSCSNMAI----- 745
QY 1156 RMDVNV-----CTCKAGYVD 1171
DB 746 -CRAVNVRYVCTCKPQFTGN 764

```

```

RESULT 9
ID 09TV02 PRELIMINARY: PRT: 1664 AA.
AC 09TV02:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Y64GI0A.7.
GN Y64GI0A.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mortimore B.J.;
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
  investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Ainscough R.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.

```

DR EMBL: AL117206; CAB60454.1; -  
 DR EMBL: AL110498; CAB60454.1; JOINED.  
 DR EMBL: AL110498; CAB57911.1; -  
 DR EMBL: AL117206; CAB57911.1; JOINED.  
 DR HSP: P00736; IAPQ.  
 DR InterPro: IPR00152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR002049; laminin\_EGF.  
 DR Pfam: PF00008; EGF\_25.  
 DR PRINTS: PR00011; EGF\_LAMININ.  
 DR SMART: SM00179; EGF\_CA; 4.  
 DR SMART: SM00001; EGF\_like; 18.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 4.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_22.  
 DR PROSITE: PS01186; EGF\_2; 24.  
 DR PROSITE: PS01187; EGF\_CA; 3.  
 DR Calcium-binding; EGF-like domain; glycoprotein; Hydroxylation; Repeat.  
 KW SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;

Query Match 6.88; Score 536; DB 5; Length 1664;  
 Best Local Similarity 21.08; Pred. No. 8; Le-36;  
 Matches 305; Conservative 142; Mismatches 451; Indels 556; Gaps 84;

QY 23 IHYNLASAIESADAYTFVFNNEAIENYIR-----EKKATSLKEDILRHVHVGELKLNKD 78  
 DB 33 IFFELGELTEIKD-----FLPFPFHARKYLRFARSRGCS-KCCLLRVANCSADL----- 83  
 QY 79 LHNGMHRETMGLFSYLAFLRNDOLYNEAPINYN-----VATDKGVIHGLEK 128  
 DB 84 CHNG-----CTCVPSEH-----NDNEQVCECPVGTGAKQCYDANECSMANGGCEHEVN 133  
 QY 129 VLEIQRKNC-----DNNDIIVRGEGKCSQAP-----CLETKRPLRET 168  
 DB 134 TIGTYRCRWMPGFELSGDGNFTCSIDECAYNSGGCDRCVSPGFRCDPSDLYLHADG 193  
 QY 169 RCTIYSIYFMKRSYFI-----GCOPQCVRIITRACMLASLAHNAKAPRGVFK 217  
 DB 194 RTCCSGHFH--ENLLIKVITSCSTDNGGCEHC-----ENDSNGEFY 234  
 QY 218 MCAIATASVMDVNGTGTCCGGLGF---NCTAC---ETCTEGKYGJHCDQASCVRGC 270  
 DB 235 -----RCRCRVGFELSENRKSCOPVDPCEFNKGG--CQHHCTNNHR- 274  
 QY 271 SQGLBGSCCCDVGWGVKCDME--ITTDNC---NGTCHISANCLDLPDCKASCACAG 325  
 DB 275 -----AQCQCYPGFH--LSYDRSCVDIDECANNGCEHFCENV---KGTYRCRKEG 322  
 QY 326 FR--GNGTVC--TAINACETSNCGSTKADKRTTPGNRVYCKAGY--TGQIVCLEINP 380  
 DB 323 YOLGRDRGTCEMLGCGGVNGGC--QHDYDOPDGGHVCRCRNCYILANDOKLCHIDINE 380  
 QY 381 CLENHGGCDRNAECTONGPNOA---VCNCLP--KYTGDKVCSLINWCLTNNGGCSPPAF 435  
 DB 381 CHENNG-----DCSQCIVNLAGSVECCKPGFRLMKDRKTCEDISESSNNGGE--QI 432  
 QY 436 CNYTEODORITCKPDY--TDDGIYCRGSITGELPKPNSTQYFFOLOEHAVER--LAGPG 492  
 DB 433 CS--MOEGGYMCSCEPGEFELSEDGHC-----HDMNECLLNG 468  
 QY 493 PFTVFAPLSSSFNHEPRIKMDQGLMSQVLRHYVGGCOLLDNLKATTSATTLQGPV 552  
 DB 469 -----GCAQ--LCKNRKSGSRCCPAGT--I 490  
 QY 553 SISVSQDTVFINNEAKVLSDDISTNGV--IHVIDKLSPNLLITPKDALGRVLDNLTT 610  
 DB 491 LAHDEKSCVAASDADIFSDNIDYSKVPJGDSIDEVIS--SIESYPADSRPL----- 543  
 QY 611 VAANHGYTKFKLLQDSLSLVIITDSITHTPVTWPTTKALEALPPEQDPLFNODNKDK 670  
 DB 544 VEGRRRHVAKACVNFQGLTSLFESSEVRTD-----PSEKCPNGF----- 582

QY 671 LKSYLKFHIVIRDSKALASDLPRSAWMTLQSGELSRGCTGSD--IGELFLNEOMCRFIHR 729  
 DB 583 -----FSTCOLSCSDQNGKCKSMRSGSLSKDCDFSGYTGEC--EQICRNGW 631  
 QY 730 GLLEPDVGAAYGIDCLMNPITLG--RCDFETTFDIPGEC--GSC---IPTPCPLKSK--- 780  
 DB 632 G-----VDCAHKCSCKLCPSTGSCRED-----PEKSSDPCPDGFGSGCCLNLCRMDC 681  
 QY 781 -----PKYKKCKIYNPLF-----RRNVCCONL---CTIV 809  
 DB 682 PNGRCDPVFGYCTCPDGLYGGSCENKCPHFTFGKNCRPPCKARENSECCDEITGKC--- 738  
 QY 810 IOTPRCHGYFMPDQ-----ACPG-----P 831  
 DB 739 -----RCKPGTYGHNCKRMRCSGLGAGCCKMSCPAGIICDPVYGDCTKKCPAGYQCNLC 794  
 QY 832 DTP-----CNNGKCRDLYTPM----- 848  
 DB 795 DQPCPAGYFGYDCEQKSCADAVASPHKSKVCHNYTGCTCLPCKTGPLCDGSCAPNTYGP 854  
 QY 849 -----GQCLCHTFNGTAC--ELCMHGRGPPD--QPRSCSEHGQDEG 888  
 DB 855 NCAHTCSCVNGAKCDESDGCHCTPGFYGATCSSEVCPGRFRGIDCMOLCKQNGAICD-- 912  
 QY 889 ITSGGECLETFGWTAAASCDPTFAVAVCTPA-----CSVHAFTTE-----NNTCYCNL 936  
 DB 913 -TSNSCCACAPGWSKCD-----KACAPGTFGKDCSKKCDCCADGMHCDPSDGECTCP 965  
 QY 937 NYEGDGI--TCTVDFCKONNGGCAKVAKCSOKT---QVSCSKKQKIDYSCIEID 990  
 DB 966 GKHKRCDCTDSGLF-----GAGCKGICSCQNGATCDSVTGSCBRRPWRCK--KCDR-- 1017  
 QY 991 PCADGVNG--GCH-----EAT--CR---MTG-----GKH 1014  
 DB 1018 PCPDGRFEGGCAIICDCTTNDTSMYNPVARCDHVTGECRCRPAWGTCPDQOTSCPLGRH 1077  
 QY 1015 K-----CECKSHYVGDGVNCE--PEOLPLDRCLD-----DNQGH 1047  
 DB 1078 GEGCHSCQCSNGASCDBVNTGFCDCPSGFMKNCSECEPBEGLMGSNCMKHCLCHMGEGCN 1137  
 QY 1048 P--DASCADLYPDPTTVGFHLSPLOGYKLTEDKAEKAEKATATYATYNOISYAKA 1104  
 DB 1138 KENGCECICIDGM-----TGPSCEFLCPFGQFGRNCAORC---NCKNGASCDRKT 1183  
 QY 1105 KYHLCSAGWLESGRAYTTTASOKCGANVGYIDYGRASKEBMDVFCYRKMDVNTC 1164  
 DB 1184 GRCECLPGW--SGE-----HCEKSC--VSG--HYGAKCEET-----CEC 1216  
 QY 1165 KAGYVGDGFS--CS 1176  
 DB 1217 ENGALCDPISGHCS 1230

RESULT 10  
 ID 09WU9 PRELIMINARY: PRT: 2906 AA.  
 AC 09WU9;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE FIBRILIN-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99350231; PubMed=10419698;  
 RA Yang Q., Ota K., Tian Y., Kumar A., Wada J., Kashihara N., Wallner E.,  
 RA "Cloning of rat fibrillin-2 cDNA and its role in branching  
 RT morphogenesis of embryonic lung.";  
 RL Dev. Biol. 212:229-242(1999).

DR EMBL: AF135060; AAD34439.1; -  
DR HSBP; P35553; IEMN.  
DR InterPro: IPR002086; Aldehyde\_dehydr.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF\_11.  
DR InterPro: IPR002212; TB.  
DR Pfam: PF00008; EGF; 46.  
DR Pfam: PF00683; TB; 9.  
DR PRINTS: PRO0010; EGFBL00.  
DR SMART: SM00179; EGF\_CA; 42.  
DR SMART: SM00001; EGF\_Like; 4.  
DR PROSITE: PS00070; ALDEHYDE\_DEHYD\_CYS; UNKNOWN\_1.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 43.  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
DR PROSITE: PS01186; EGF\_2; 36.  
DR PROSITE: PS01187; EGF\_CA; 43.  
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
KW SEQUENCE 2906 AA; 313371 MW; 9EB64E727044EF58 CRC64;

Query Match 6.6% Score 519.5; DB 11; Length 2906;

Best Local Similarity 24.1% Pred. No. 4.9e-34; Mismatches 407; Indels 419; Gaps 87;

Matches 301; Conservative 124; Mismatches 407; Indels 419; Gaps 87;  
QY 190 QCVFTITIRACMLASLANAKPAPG---EVRMCALGTASVWDVNGTGT--COCGLGFNG 244  
Db 501 RCITPVSYSRC-ECNMGYK-QDANGDCIDVDECTSNPCSHQDCVTPSYCKKAGNR 558  
QY 245 T---ACCTCTEGKYGIMHCO-ACSCVHRCSSQGPLDGS--CQCDVGMRGKCDMETT 297  
Db 559 TPTKQACIDIDE---CIQNGVLCNKRGVN---TDSFQCIQCNAGF---ELTT 602  
QY 298 D--KNG--TCFTSANCL---LDPDGKASCKCAAG--RGNGVCTAINCET---SN 343  
Db 603 DGEKCVGDECTTINMCLNGKINEDGSFKCVKRGFVLARNGRCTTIDECQITGICM 662  
QY 344 GGC---STRADCKRTTPG---NRVCV----- 363  
Db 663 GCHINNESGFRDC---PPGLAVGVDRGVDTIMRSTCYGEIKKGVCPVPPGAVTKYE 719  
QY 364----- 363  
Db 720 CCCANPDYGFGEPCQPCPAKNSAEFHGLCSGVGITTVDGRDINECALDPDICANGICENL 779  
QY 364-----CKAGYTG-D-GIVCLEINPCLENHGGCDRNAECTQTGPNQAVCNCLPKYT--G 412  
Db 780 RGSYRCNCSGTEPDASRNCIDDECLVNRLLCD-KGLCRNT-PGISTCTCPPEYFRT 837  
QY 413 DGKVCSLINVLGN---NGGSPFAFCVNTBODORICTCKP--DYTGDIYCRGSIVGEL 467  
Db 838 ETETCEDEVNECESNPVCVAGACR-----NNLGSFHCECSPGSKASSTGLICISGLKTC 890  
QY 468 PKNPSTSGYFQLOEHAHVAELAGPEPTVVPAPLSSSFHHEPRIKDMDQGLMSOVLRYHV 527  
Db 891 WLNIQDNCEVNINGATLKS-----ECCATLGTFA-----WGSF----- 923  
QY 528 VGCQOULLDNKLVTSATTLGEPYSIVSODYVFINNEAVL-----SSDIISTNGVIA 582  
Db 924 --CERCELD-----AACPRGFAIRIKYCTCEDV---NECEVPEVCNNGRCVNNKSGSFH 971  
QY 583 VIDKLSPKLLITPKALIGRVLQNLTTVAANHGYTKFSKLIDPSGLSV-----ITDSIH 638  
Db 972 ---CEPEGLTL--DGTGVCLD--VRMEHCFKMD--EDBCVAPVPGKFRMDACC 1018  
QY 639 TPVYVFWPTD-----KALEALPPEQOQFLFNQDKDKLSTLKHVIRDSALASD 689  
Db 1019 CAVAAAMGTCECECPKPKTEYETLCPRGPGF---ANGDILITGRPFYKDIINECALPQM 1075  
QY 690 LPRSAWMTLGGSELVSCGTG-----SDIGELFLNEMOKRFTHRGILPVGAVY 739  
Db 1076 CTYQKCRNTI--GSFKCNCNSGFALDMEERNCTDIDECRISPDLC---GNGICVNTPGSF 1130

QY 740 GIDC-----LIM-----NPTL--GGRC-DFTFTFDIPGEGCSCTTPKCP 777  
Db 1131 ECECEGEYESGFMMKNCMDIDECERNPLLCRGCTVTEGSF-----QC-----DCPL 1179  
QY 778 KSKPKGVKKKCI-YNPLFRNRVVECCQ-LCTVVIQTPRC-CH-GY-TPMPCQACP----- 828  
Db 1180 GHELSPSREDCIDINECSLDL--CRNGKCVNMIGTVQCSQCNPGYQATPRQGSIDIDE 1237  
QY 829 ---GCPDTPCNRNGMCRDLYTPMGQCLHTGF-----NGTAC---ELCWHGREGPDQP 876  
Db 1238 CMINMGCDTQCTN---SEGSY-----ECSSEGYALMDGASCADIDC-----ENNP 1283  
QY 877 RSCSEHGQCDGEGTSGGELCTETGWTASCDTPAV-FAVC--TPACSVHATCTENN--- 930  
Db 1284 DIC-DGGQC-TNIPGEYRCLYDGF-MASMDMKCTIDVNECDLNNICMFGBC-ENTKGS 1339  
QY 931 -TCVGNLWYE-GDGIT-CTVYDFCKQNNGGCAKAVAKCSQKQTVSCSKKKGIGYSCI 987  
Db 1340 FICHQQLGYSVKKGATGCTDVEDEIGAHNCDMAHASCINVRGSEKSCREGWVGKICI 1399  
QY 988 EIDPCADGVNCGCHEHATCRMTGPKHKCECKSHYVGDVCEPRLPLDR-----LQD 1042  
Db 1400 DIDECANQTH-QCSINACQVNT-POSTACACSEGTGQDGFICS---DYDECAENINLCE 1453  
QY 1043 NGQC-----HPDASCADL---YFQDTTV-----GVFHLRSPLGQ 1073  
Db 1454 NGQCLNVGAYRCECEMGFTPASDSRQDIDECSEFQNICVFQTCNNLPGMFHCICDDG- 1512  
QY 1074 YKLTFDKAKEACA---KKAATATATYNOLSTYAKAKYHILCSAGMLESGVAYPTTYASKC 1130  
Db 1513 YGL--DRTGHCITDIDECADPLINCYNGLCVNTPGRYE-CNC-----PPDE--QL 1556  
QY 1131 GANYGVIVDYCSRANKSEMDVFCYRMK-----DVNCTCKAGYVGDGSCS 1176  
Db 1557 NATGVCVD--NRVN-----CY-LKFPGRGDSLSCKT-ENGVGVSCS 1596  
RESULT 11  
ID 025253 PRELIMINARY; PRT: 2653 AA.  
AC 025253;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE NOTCH HOMOLOG SCALLOPED WINGS (SCL).  
GN SCL.  
OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Oestroidea; Calliphoridae; Lucilia.  
OX NCBI\_Taxid=7375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SS SEEKING; PubMed=8807304;  
RX MEDLINE=96400928; PubMed=8807304;  
RA Davies A.G., Game A.Y., Chen Z., Williams T.J., Goodall S., Yen J.L.,  
RA McKenzie J.A., Batterham P.;  
RT "Scalloped wings is the Lucilia cuprina Notch homologue and a  
RT candidate for the modifier of fitness and asymmetry of diazot  
RT resistance".  
RL Genetics 143:1321-1337(1996).  
RN [2]  
RP SEQUENCE OF 39-265 FROM N.A.  
RC STRAIN=SS SEEKING;  
RA Chen Z., Newsome T., McKenzie J.A., Batterham P.;  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 39-265 FROM N.A.  
RC STRAIN=SS SEEKING;  
RA Chen Z., McKenzie J.A., Batterham P.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U58977; AAC36151.1; -.

DR EMBL: AF032672: AAC36152.1; -  
 DR EMBL: AF032670: AAC36152.1; JOINED.  
 DR EMBL: AF032671: AAC36152.1; JOINED.  
 DR EMBL: AF032673: AAC36153.1; -  
 DR HSSP: P00740: 1EDM.  
 DR InterPro: IPR002110: ANK.  
 DR InterPro: IPR000152: Asx\_hydroxyl.  
 DR InterPro: IPR000561: EGF-like.  
 DR InterPro: IPR000742: EGF-2.  
 DR InterPro: IPR001881: EGF-Ca.  
 DR InterPro: IPR001438: EGF-II.  
 DR InterPro: IPR000800: Notch.  
 DR Pfam: PF00023: ank; 6.  
 DR Pfam: PF00008: EGF; 36.  
 DR Pfam: PF00066: notch; 3.  
 DR PRINTS: PRO0010: EGFBL00D.  
 DR PRINTS: PRO1452: NOTCH.  
 DR SMART: SM00248: ANK; 4.  
 DR SMART: SM00179: EGF\_CA; 24.  
 DR SMART: SM00001: EGF\_Like; 11.  
 DR PROSITE: PS00088: ANK\_REPEAT; 5.  
 DR PROSITE: PS00297: ANK\_REPEAT; 1.  
 DR PROSITE: PS00010: ASX\_HYDROXYL; 22.  
 DR PROSITE: PS00022: EGF\_1; UNKNOWN\_34.  
 DR PROSITE: PS01186: EGF\_2; 28.  
 DR PROSITE: PS01187: EGF\_CA; 21.  
 DR ANK repeat: Calcium-binding; EGF-like domain; Glycoprotein;  
 KM Hydroxylation: Repeat.  
 SQ SEQUENCE 2653 AA: 285928 MW: 66A2A058FE6C329 CRC64:

Query Match 6.6%; Score 519; DB 5; Length 2653;  
 Best local Similarity 22.7%; Pred. No. 4,7e-34;  
 Matches 298; Conservative 117; Mismatches 416; Indels 482; Gaps 84;

137 CDNDNTIIVGEGCKGSAQAPLETRKRLRETRKCIYSIFEMCK--RSVFICQPO----- 190  
 DB -203 CDNN-----PCKHNGTGC-LNT---HGSYQCMCPAGYTGKNCESKYPSPCQCN 248  
 QY 191 ---CVRTIITRACML-----ASLAHNKAPRGEVKKKALGTASWMDGVN----- 231  
 DB 249 GGTGRTSTGLIYECKCPREGYGGKNCQENIDDCPGH--LCQMGGTGI-DGINSYHCACPPNY 305  
 QY 232 -----GTGTCQCGLGFNGTACE----- 248  
 DB 306 TGENCEKDVDECAIRPSVCONGATCTNSGYSICVNGMTGPCCSINIDDCIAAAGFYG 365  
 QY 249 -----TCTEGKYGI--HCDQACS--C-VHRCSSQGL-GDGSQDCDVGMRGV 289  
 DB 366 ATCIDVGVSFYCRCTPGKTLGLLHDADCTSNPCHADALCDTSPINGSYTCPCATGYKGV 425  
 QY 290 KCDMEITTDNCN--GTCHTSANCLLDPDGKASCKCAAGFNGNIVC-TAINACTSNGGC 346  
 DB 426 DCSEDI--DECIDQSPCEHNGVGVNTP-GSFRNCSSQGF--TGPRCTINIECESH--PC 478  
 QY 347 STRADCKRTTPGRRVGVCKAGYTG-----DGICYLEINPCTLENHGGCDRNAECTQTGPNQ 401  
 DB 479 QNNGSGC-LDDPGTFRCVCMGCTGTGCEIDINEC-QSNPCL-NGICNDMI-----NG 528  
 QY 402 AYCNCLEPKYTGDKVSL-INVLT-----NNGGCSPPAFNGYTEQDORICTCKRDTYGDG 456  
 DB 529 FKSCSALGFSGSR--COINIDDCQSPCRNNGICRD-SIAGT-----CQCPRGYT--G 577  
 QY 457 IVRGSIYGLGPKNPSTISQVFFOLOEHAVALAGPGFT--VFAPLSSFNHPRIKAMDQ 515  
 DB 578 LSCSEINI-NMCNSNPC-----HRGKCIDGDNRFCTVCPDGFGLYLCOTQDINECE- 625  
 QY 516 OGIMSGVQLRY--HYVGGQQLLDNLKVTTSATTLQGEPPVSISSQDVPFINNEKAVLSSD 573  
 DB 626 ----SNPCQVGGGCHVDKRVGVMCHCLAGTS-----GKDCETINNE----- 661  
 QY 574 IISTNGVIHVIDKLSPKLLITPKDALGRVLONLTVAANHGYTKFSKLIODSGLLSVI 633

DB 662 -----CH-----SNPCNNKATCIDGI-----NKYTQCVPGFT----- 689  
 QY 634 TDSIHTFYVYVWPTDKALEALPRPEQOQFLFNQDDKKDLKSLTKRHYIRDSALASDLPRS 693  
 DB 690 --GVHCEINI-----NECASNPANNGVCMDLVNGYKCECPGRFY---DPKCL-SDVDEC 738  
 QY 694 ASWKTLOG-----SELSVRCGTG-----SPIGELFLEQKCRFTIHRLLDVGAVY 739  
 DB 739 ASNPICINGRCEDGINEFICHPPGYGGKRCENIDECSSNP--CQ--HGFCYDEINAF 794  
 QY 740 GIDCLLNPTLGRCDFTFTFDIPGEC--GSCI-----FTPK-CPLKSKPK 782  
 DB 795 KQCG--MPQYGLKCEINIDDCINNPCANGSTCIDKYNKYKVCYKVTYTGDDCSKLDPC 852  
 QY 783 GVKR-----KCIYNP--LPR-----RNEGCONL----- 805  
 DB 853 ATNRCRNDAKCTPSNPLDPSCTCKLGTYGRYCEDEIDECKLSTPCRNNGATCHVVPGSYR 912  
 QY 806 CTVAIQTPRCHGYFMPDCQACPGPDT-----PCNNRGMCRD--LYTPMGQCLCHTFGN 858  
 DB 913 CI-----CAKYEIGHDCAI--NTDDCAFPQONGGTCLDGIQDYT-----CLCVPGFD 958  
 QY 859 GTACEL-----CMHGRF-----GPDQCPRSCSEHGQC 885  
 DB 959 GKHCETDINECLSMPCQNGATCRQYVNSYTCPCPLGFSGINCQTNDEBDCTESSCMNGTIC 1018  
 QY 886 DEGITGSGECLCEFGMTAASCDPTFAVAVCTPACSVHATCTEN--TCVCNINTYGSG 942  
 DB 1019 IDGI-NSYNCSCLPFGYGSNQ--YKINKDSQPCQNGATCHEGDEYTCGSGYTGK- 1074  
 QY 943 ITCT-VDFCKONNGGCAKAYAKCSOKGTQVSCSKCKGKGBGYSC-LEIPRCAD-----G 995  
 DB 1075 -QCTIDYDWDCKTS--PEENGATCTQVKNQNSCKRAPWTGK--LCDVEMVSCSDAALRK 1129  
 QY 996 V-----NGCCHENATCRMTGPKHKCECKSHYVG----- 1024  
 DB 1130 VSLEFOLCNNGTCKEHNMI-----HRCYKQGGTGYSCQGEINECESQPLANGTCRDL 1182  
 QY 1025 -----DGVCEPEPDLDRG-----LDNNGCCHPDASCADLYFDOTTYGVNHLR 1068  
 DB 1183 IGSYACVCRKRGFGQONE--LNLIDCSPNPCQNGGTCH-----DL-----VNTFSCS 1227  
 QY 1069 SPLQOYKLTFFDKAKECAKAEATATATYNOLSYAQKAYHLCASAGWLSGRVAYPTTASQ 1128  
 DB 1228 CPKCTACILCEVNEINDCKRSGC--HNGTCTIDRVGGFEC-----ACPPGVGS 1273  
 QY 1129 KCGANV-----VGIVDGSRANKSEMDVFCYRM-KDVNCTCKAGYVG 1170  
 DB 1274 RCEGDINECLSNPCSNAGTLD-----CVQVLNVNHYCNCKKPGYMG 1312

RESULT 12  
 Q9W0H8 PRELIMINARY: PRT: 2872 AA.  
 ID Q9W0H8;  
 AC Q9W0H8;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE FIBRILLIN-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90032689; PubMed=9815129;  
 RA Kanwar Y.S., Ota K., Yang Q., Kumar A., Wada J., Kashihara N.,  
 RT "Isolation of rat fibrillin-1 cDNA and its relevance in metaphoric  
 RL development.";  
 Am. J. Physiol. 275:F710-F723(1998).  
 RN [2]

RP SEQUENCE FROM N.A.  
RA Kanwar Y.S.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF135059; AAD34438.1; -  
DR HSSP: P35555; 1ARI  
DR InterPro: IPR00152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR002212; TB.  
DR InterPro: IPR000823; ZnF-C2H2.  
DR Pfam: PF00008; EGF\_45.  
DR Pfam: PF00683; TB; 9.  
DR SMART: SM00179; EGF\_CA; 41.  
DR SMART: SM00001; EGF\_like; 5.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 42.  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
DR PROSITE: PS01166; EGF\_2; 38.  
DR PROSITE: PS01187; EGF\_CA; 41.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.  
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
KW SEQUENCE 2872 AA; 312069 MW; 0CAFE73B87A80280 CRC64;

Query Match 6.48; Score 499.5; DB 11; Length 2872;  
Best Local Similarity 21.88; Pred. No. 2,4e-32;  
Matches 251; Conservative 111; Mismatches 322; Indels 469; Gaps 67;

QY 189 POCVRIITTRACWLAS-LAHNAK--PAPG-----EYKMCALGTSV 226  
DB 443 PKVLPEVNTDYCQLVRYLCOGRCTIPRGSYRCENKNGFOLDINGECIDVECEKNPTG 502  
QY 227 WDGVNCTG--TCQCGLGFNGTACTCTEGKYGICDQ-ACSCVHGRSGQPLGDS--CD 281  
DB 503 GECINNGSYTCHCRAGYQSTL--TRTECRDIDECIDNGRCNNGRCLN--TGSSFCV 557  
QY 282 CDVGRW---GVKCDMETTDNC-----NGTCHTSANCLDIPDGKASCAAGFR--G 328  
DB 558 CNAGFHVTRDGRKNCE---DMDECSIRNMCCLNGMC-----INEDGSFKCTCKPGFOLAS 607  
QY 329 NGVTCATINACET---SNGCC-----KRTTGNRV-- 361  
DB 608 DGRCKDINECEETPGICMNGRCVNTDSYRCEPGLAVGLDGRVCVDTMHRSTCYGYR 667  
QY 347 -----STRADC-----KRTTGNRV-- 361  
DB 668 RGCCVCKPLFCAVTRKSECCASTEYAFGEACOPCPAONSAYEQALCSSGSPRTAGSDINE 727  
QY 362 -----CVCKAGYTGD--GIVCLEINPCLENNHGGCDRNABCTQYG 398  
DB 728 CALDPDICPNGICENLRGTYKICNSGEVDITGKNCVDINECVLNSLLCD--NGCCRT-- 785  
QY 399 PNOAVCNCLPK---YTGDGKVCSLINCLTN--NGGC--SPFAFCVNTEDQRICTCKP 450  
DB 786 PGSVVCTC-FKGFYRKDLKTCEIDECSSPCINGVCKNSPSGF-----ICEESP 835  
QY 451 DYTGD--GIVCGRGSIYELPKNPSTSQYFQLOEHAVERELAGPFTVFAFLSSFNHP 508  
DB 836 ESTLDPTKICIEITIKGT----- 853  
QY 509 RIKMDQOGLMSQVLRHVHVGCCOLLID-NLKVTSATTLQ-----GEPSV-- 554  
DB 854 -----CMQTVIDGRCEININATLTKSECCSSIGAMGSPCTIQ 892  
QY 555 -----SVSODTFVINEAKVLSDDIISTNGV--HVIDKLSPNLLITP 597  
DB 893 VDPICGKGRKIKGTGCEDI-----NECEVFRG--VCKNGCLVNSRGSFKCCPSGMLT-- 944  
QY 598 KDALGRVQLNLTVAANHGYTKFSKLIDQSLSVT-----TDSIHTPTVYFWPTD----- 648  
DB 945 -DATGRICLDIRL-----ETCFLLKYDDECELTPIAGRHNRDACCSSVGAAMGEEDEEC 997  
QY 649 -----KALEALPREQDFLRNODKKDLKSLYLFKHNVRDSEALASDLPRASMKWTLOGSE 703

DB 998 PLRNSREYELCPKPGPFA---TKDITNGKPFKDIINECKMIDSLCTHGCRCRTI--GS 1051  
QY 704 LSVRCGTSGDIDGELFLINOMCRFLIHRGLLPVGAAYGIDCLLMNPTLGRCDTFTFDIP 763  
DB 1052 FKCRDSDGFALDS---ERRNCT-----DIDECRISPDLCR----- 1084  
QY 764 GEGSCIFTP-----KCPYKSKPRG--VKKKCIYNPLPFRNRVEGQNLCTVVIQTPRC 816  
DB 1085 ---GQCWTPGDPDEKCD-EEYESGFMKMKM-----DIDECQ-----RDPLLC 1125  
QY 817 HGTFMPDCQA-----CPGGRPD-TRCNNRGRCDL-----YTPMGCL-----C 853  
DB 1126 RGGICHNTEGSRCEPCBGHQLSP--NISACIDINECELSANLCPHRCVNLIGKYECAC 1183  
QY 854 HTGFNGT-----ACE-LCW-H--GRECPDOP-----RSCSEHGOC 885  
DB 1184 NPGYHPTHDLFCVYDIDECSTLMNGCEFCFINSDSGYSQCPGAFALMPDRSCTDIDEC 1243  
QY 886 DE-----GITSGECLCETGWTAA-----SCDPTPAVFAVCTPACSVHA 924  
DB 1244 EDNPNICDGGCCTNLPGEYRCLCYDGFMASEDMTQVDVNECD-----LNPNICLSG 1295  
QY 925 TCTERN-----TCVCNLNTEG-DGIT-CTVDPFCQONNGCKAKVAKSGKQVSCSKKG 978  
DB 1296 TC-ENTKGSFICHDDMGYSKKKGTGCTDINECEIGAHNCRHVACTPTAGSFRKSCSPG 1354  
QY 979 YKGDGSCIEIDPCADGVNGGCHENATCRMTGPKHKCEKSHVVGDDVDCPEPQLPLDR 1038  
DB 1335 WIGDGICTDIDECSSNGTH-MCSQHADCKNT-MQSYRCLCKDGTGGDFTC-----TDLDE 1408  
QY 1039 C-----LQDNGOC 1046  
DB 1409 CSENLNLSNGOC 1421

RESULT 13  
ID 035516 PRELIMINARY; PRT; 2470 AA.  
AC 035516;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CELL SURFACE PROTEIN.  
GN NOTCH2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57B/6; TISSUE=THYMUS;  
RX MEDLINE=93178563; PubMed=8440332;  
RA Lardelli M., Lendahl U.;  
RT "Notch A and Notch B - two mouse Notch homologues coexpressed in a  
RT wide variety of tissues."  
RL Exp. Cell Res. 204:364-372(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57B/6; TISSUE=THYMUS;  
RA Hamada Y., Higuchi M., Tsujimoto Y.;  
RT "Complete amino acid sequence and multiform transcripts encoded by a  
RT single copy of mouse Notch2 gene."  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL: D32210; BAA22094.1; -  
DR HSSP: P16109; 1FSB.  
DR MGD: MGI:97364; Notch2.  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000742; EGF\_2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF\_II.  
DR InterPro: IPR000800; Notch.



DR Pfam; PF00023; ank; 6.  
 DR Pfam; PF00008; EGF; 35.  
 DR Pfam; PF00066; notch; 2.  
 DR PRINTS; PR00010; EGFBLD.  
 DR PRINTS; PR01452; NOTCH.  
 DR SMART; SM00248; ANK; 4.  
 DR SMART; SM00179; EGF\_CA; 22.  
 DR SMART; SM00001; EGF\_Like; 12.  
 DR SMART; SM00004; NL; 3.  
 DR PROSITE; PS50088; ANK\_REPEAT; 4.  
 DR PROSITE; PS50297; ANK\_REPEAT; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_33.  
 DR PROSITE; PS01186; EGF\_2; 27.  
 DR PROSITE; PS01187; EGF\_CA; 22.  
 DR ANK repeat: Calcium-binding; EGF-like domain; glycoprotein;  
 KW Hydroxylation: Repeat  
 SQ SEQUENCE 2470 AA; 265325 MW; B55A31B35242716D CRC64;

Query Match 6.2%; Score 489; DB 11; Length 2470;  
 Best Local Similarity 22.7%; Pred. No. 1,5e-31;  
 Matches 282; Conservative 119; Mismatches 445; Indels 398; Gaps 78;

OY 132 IQKNRCNDNDT-----IIVRGEGC-KCSQAAPCPLETKPLRETRKCTYSIFYMGR 180  
 DB 302 LQPNACOMGCTCTNRNNGYGVGVNGSGDDCSENI-----DCAYV----- 343  
 OY 181 RSVFIGCQPCVCRITITRACWLASLAHNAKPARGEVKKMCALGTASY-----WDG 229  
 DB 344 -----SCTPG--STCIDRVSFSCLPREGKAG-----LCHLDACISNCHKALGALDTNP 392  
 OY 230 VNGTGCOCGLGFNCTACPTCEGKYGIHQDQCS--CVH--GRSGGPLGDS--CDDCV 284  
 DB 393 LINGYITCTPCQYKAKADTEDVD-----ECAMANSNPGCHACKVNA--TDAPHCBLK 444  
 OY 285 GMRGVKCDMEITTDNCG-TCHTSANCLLDPDGKASCACAGFRNGTVC-TAINACET- 341  
 DB 445 GVAGRCRMDI--NECHSDPCQNDATG-LDKIGFTCLCMPEFK--GVNCELEVMEGCSN 499  
 OY 342 ---SNGGSTKADCKRTTPGRNV-CVCKAGTYGDGIVC-LEINPCLNHHGGCDRNAECTO 396  
 DB 500 PCVNNGQGVCKV-----NRFQCLCPGGFTG--VCOIDIDDC--SSTPLMNAKCID 547  
 OY 397 TGPNAVNCNCLPYTGDDGVCSLIWVCLTNNGCSPPAFACNTE---QDQITCCKPY 452  
 DB 548 -HPNGYECOCATGFTG-----ILCDENIDNCPDP-CHHGQCCODGIDSTICTONPEY 597  
 OY 453 TGDGIVCRGSIYGLPKNPSTSQYFFLOEHAHVRELAGPGPTVPAPLSSFNHEPRID 512  
 DB 598 M--GAIGSDQI-DECYSSPC-----LMDKCIDLVN-GYCCNCPGTSGLCEINFPD 646  
 OY 513 WDOGLMSOVL-----RHVVGCCQQLLDNLKVTTSATTLGCEPVSISVSQDTVEINNEA 567  
 DB 647 CASNPCHMGVYCDGINRYSV-----CSPGFTGQRNIDIDE---CASNPC 689  
 OY 568 KVLSDDIITNCGVHVHIDKLSPKMLITPKDALGRVILQNLTTVAANHYTEFSKLIDS 627  
 DB 690 RKGATCINDVNGRCICTCEGPHHPSYCOVNECLSNPCIHGCTGGLSGY---KCLDA 745  
 OY 628 GLLSYITDSIHPTVVFVPTDKALEALPPEODFLFNODNKKLSYLKFHVIRSKALA 687  
 DB 746 GMVGVNCE-----VDKNECLSNPCQNGTCN--NLVNGY----- 777  
 OY 668 SDLPASAKMTIQSELSYKRGCTSDIGELFLNEQMRPIHRLGLFDVGAVAGIDCLLMN 747  
 DB 778 ---RCTCKKGGKYNCOV-----NIDECA SNP--C--LNGCTCFDDVSGYTCMCL-- 821  
 OY 748 PTLGRCQPTFTFDIPREGC-----SCIFTP-----KCPKX-----SKP-- 781  
 DB 822 PTTGNCQOTVLAFCSPNCEANAACKEAPNENPESFSCCLAPGMOGKRCYDVDECSKPCM 881  
 OY 782 -KGV-----KKCIYNPLPFRNVEGCON--LCTVVIQTPR 814

DB 882 ANGVCHNTGGSYVCEPCPCFSGMDCEDINDCLAMP-----CQNGSGCVDHVNFPS 932  
 OY 815 C-CHGYFMRD-CQA-----CPGSPDTPCNNRGMCRDLTYPMGQCLCHTGFNGTACE----- 863  
 DB 933 COCHPFGTGDCKQDTMNEC---LSEPCKNKGTCSD-YVNSYCTCPAGHGVCHENNIDE 988  
 OY 864 -----LCWHGRFQP-----DCQPRSCSEHGQCDGITGSGEC 895  
 DB 988 CTRESSCFNGCTGVDDGINSFCLCPYGFPGFCLHIDNECSSNPCLNAGTCVGL-GTYRC 1047  
 OY 896 LCEGTWTAASCDTPPAVFAVCTPA-CVSHATCTENNT---CVCNLYNESDGIYCTVDF- 950  
 DB 1048 ICPLEGTGKNQCT---LVNLCSRSPCKMKKGTQVQKARPHCLCPGWM--DGAYCDVLNVS 1102  
 OY 951 CK-----QNNGGCAKAKCSQKGTQVSCSKKKYKKGQDGSICFIEDPCADGV 996  
 DB 1103 CKAALQKVPYEHLCQHSQICINA-----GNTHHQCPILGYTG-SYCEQLDFCA-- 1152  
 OY 997 NGGCHHATCRMTGPGKHKCECKSHYGVDPCEPEOLPLDRCLQDNGCHPDASCADLY 1056  
 DB 1153 SNPCQHGATCN-DFIGYRCCEVPCT--QGVNCEYE---VDEC--QNOPCOMGTCIDL- 1203  
 OY 1057 FQDTVGVFHLRSLPGQKYLTPDKAKKACAKAEATITATYMNLSYAQKAYHLSAGMLES 1116  
 DB 1204 -----VNHFKSCSPGTGRLCLCEENIDCA-----GGPHCLNGGQCV 1241  
 OY 1117 GRVATPTT-----YASQKCANVGVIVDYSRANKSEMDVFCYRMK-DYNTCKKAGYV- 1170  
 DB 1242 RIGGYTCRCLPGFAGEREGDINECL---SNPCSE-GLSDCVQLKNVNCICRSATGCR 1297  
 OY 1171 -----DGFSC-----SGNLQ 1181  
 DB 1298 HCTFLDVCPOKPCPLNGGTCVAVSNMPPGFCRCPPGSGARLQ 1341

RESULT 14  
 O9GPA5 PRELIMINARY; PRT; 2524 AA.  
 AC O9GPA5.  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PUTATIVE NOTCH RECEPTOR PROTEIN.  
 GN NOTCH.  
 OS Brachiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.  
 OX NCBI\_TaxID=7739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LARVA;  
 RA Holland L.Z., Burgtorf C., Holland N.D., Lehrach H., Tamme R.,  
 RA Abl-Rached L., Pontarotti P., Lardelli M.;  
 RT "Cloning and developmental expression of the amphioxus homologue of  
 RT Notch (Amphinoct): evolutionary conservation of multiple expression  
 RT domains in amphioxus and vertebrates";  
 RT Sumblin (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; Y12539; CAC19873.1; -;  
 DR HSSP; P00740; 1EDM.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF\_Like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_II.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR000800; Notch.  
 DR Pfam; PF00008; EGF; 36.  
 DR Pfam; PF00066; notch; 3.  
 DR Pfam; PF00010; EGFBLD.  
 DR PRINTS; PR00011; EGFAMININ.  
 DR PRINTS; PR01452; NOTCH.

DR SMART: SM00248; ANK; 6.  
 DR SMART: SM00181; EGF; 37.  
 DR SMART: SM00179; EGF\_CA; 34.  
 DR SMART: SM00001; EGF\_like; 13.  
 DR SMART: SM00004; NL; 3.  
 DR PROSITE: PS50088; ANK\_REPEAT; 4.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR ANK repeat: EGF-like domain; Glycoprotein; Hydroxylation; Receptor;  
 Repeat.  
 KW SEQUENCE 2524 AA; 270969 MW; C2CA57E306D23BC9 CRC64;

Query Match 6.2%; Score 486.5; DB 5; Length 2524;  
 Best Local Similarity 22.1%; Pred. No. 2,5e-31;  
 Matches 271; Conservative 107; Mismatches 413; Indels 437; Gaps 72;

QY 236 CCGGAGFNCTACETCTEGKYYGHCQACSCVHG-----RCSCG 273  
 DB 233 CHMSGFTGDDDEVVDDCVDLHCENGAACVDGVNEYTCTCPSONAGRYCNEDVDECMQS 292  
 QY 274 P-----LGDGSCDGVNGRYKCDMEITTDNC-NGTCHTSANCLDPPDGKASC 320  
 DB 293 PICTLNCGTCHNTVGGYSCVYNGMIGDCCSENF--DDCASAACEFGATC-HDRVGFEMC 349  
 QY 321 KCAAFERNGVCTA1INACETSNCGCSTKADCKRTTP--GNRVCKAKGYTGDGIYCL-E 377  
 DB 350 ECAPG--KTGLLCHLDADACESS--PCNEGALCD-TNPVNGOPICPCPDGY--EGQLCMQD 402  
 QY 378 IMPCLEHNGGCRNAECTGTGPNQAVCNCLPKYTGDKKVCSL-INVCLTNNGSCSEFPAC 436  
 DB 403 IDECALGENPCHDEGCNNV-PGSFTCTCTDGTGDR--CEVINMECASN--PCNOGTC 457  
 QY 437 NYTEDDORICTCKPYTGDIYGRSISYELPKNP-----STSOYFFOLQ----- 481  
 DB 458 -IDDIGEFACAMPFAGD--LCETDV-DECASSPCLNCLCDGINKYCECEDPREGTT 513  
 QY 482 -EHAVERELAGPPTFYFAPFLSSSFNHPRIKMDQGLMSQVLRVHVVCQQLLDNLKV 540  
 DB 514 CENNINECAN-GPCRGAHCS-----DLVT 537  
 QY 541 TTSATTLQG-----EPVISYSQDTVFINNEAKVLSDDIISTNGVT--HYDKL- 587  
 DB 538 TYACTCLEGFTGDEINIDQSNOCCHGTGVDGASFTGCEPQYCNPLCESPVEDC 597  
 QY 588 -----LSPKLLITPKDALGRVL-----QULTVVAANH 615  
 DB 598 SPCPONGTCELDVNGYRCNCLAGTSGSNCEVNODDCTNLCYHGVCOGLNDYTCCCG 657  
 QY 616 GYTK-----FSKLIQDSGLSVITDSIHTPVTVFWPTDKALEALPPEQDQFLNQ 665  
 DB 658 GYEANNCEREIDECASSPCHNGGI-----CHDLVNAE--SECPRGYHDLQCYEN 705  
 QY 666 DN-----KDKLSKY-LKREIVTRDSKALASDLPRSSA----- 695  
 DB 706 VNECESSPCAHGTCCDGINDYCTCENGEGKNCVDVNIIDECSNPNQHEGCDGIGRYE 765  
 QY 696 KRTLOGSELVRCGTGSDIGELFLNEQMCRTIHRGLLPVGAAYGIDCLMNPITLGRCD 755  
 DB 766 CCGCLGYE-GVNCIDINTD--ECASNP--CQ--NGRCILDGANNVYDCDL--PFTVGTNQ 816  
 QY 756 TTTTEDIPEGC---GSCI-----FTPKPLKSRPKGVK-----KKCIYNLPPRR 797  
 DB 817 TELAPCRPAPCENLACIPISADYQFTFCNC-----ADFEGETCADDINECSNPNC---K 868  
 QY 798 NVEGCONLCTVVIQTPRCCHGYFMPDC-----QACPGSPD---TPCNRNRCMRD--L 844  
 DB 869 NCAPCNL-----EGDFRCDCLMFAGELCSVNIIDODDPCCNGSGTCDNDGINS 917  
 QY 845 YTPMGQCLCHTGNGTACEL-----CMHGRFGPDCQPR- 877  
 DB 918 YI-----CSCMPGFGTNCEDIDICTSNPCONGGOCIDAVNGYACDCVVGFTGNTQINK 973

QY 878 -----SCSEHGCDEGTTGSGECLCTETGWTASCDTPTPAVAVCTPACSVHATCTENN- 930  
 DB 974 DDCSTSSCFSGGTCTIDG-NTFTCHCPSGFTGSMQ--HEINEDSNPNCONGATCVDDTG 1030  
 QY 931 --TCVGNLNEYEDGTTC-TVVDPCKN---NGS----- 957  
 DB 1031 YFSCICTGYE--GVVCSQOKDLCADDDPCRNNGTCTGSDRYECLCEDEMTGLIDMTKV 1088  
 QY 958 -CAKAK-----CSQKGTQVSCSKKKGYKGDGYSCLEIDPCADGVNNGCH 1001  
 DB 1089 SCMAAASGRNNSLANLNCNGGTCTVDTGNSHNCCNCAAGYNG--SYCSEIDECA---SPQ 1144  
 QY 1002 EHATCRMTGPGKHKCEKSHVYVDGVEPEQLPLDRCLQDNGQCHPDASCADLYPDOTT 1061  
 DB 1145 NAEGR-DGLGYTCACRPGY--QGVNCEOE--INECT--SNPCONGGTCTIDM----- 1190  
 QY 1062 VGVFHLRSLGQKTLFPDKAKENAKETATITYNOLSTYAKAKYH--LCSGMLESGRV 1119  
 DB 1191 VNEYRSCPPGTGGLCEITNDCNCFAGAC-----YHDTGVVDGIEFTCR 1235  
 QY 1120 APTTYASQKCANV-----VGIVDYSRANKSEMDVFCYRMK-DVNYCTCKAGY 1168  
 DB 1236 CRP-GVYGRCEGDVNECLSNPCDAEGTLD-----CVQLENDYSCDCKRXY 1280  
 QY 1169 VG-----DGESC--SGN 1178  
 DB 1281 TGRRCERTVDSCEPDLCLNGACASQSGN 1308

## RESULT 15

Q9XWD6 PRELIMINARY; PRT; 1111 AA.

Q9XWD6;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE V47H9C.4 PROTEIN (CED-1).

GN V47H9C.4 OR CED-1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN (1)

RP SEQUENCE FROM N.A.

RA Harris B.;

RL Submitted (OCR-1998) to the EMBL/Genbank/DBJ databases.

RN (2)

RP SEQUENCE FROM N.A.

RA MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,

RA Gardner M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gartner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Tarry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

RN (3)

RP SEQUENCE FROM N.A.

RA MEDLINE=21097720; PubMed=11163239;

RA Zhou Z., Hartwig E., Horvitz H.R.;

RT "CED-1 is a Transmembrane Receptor that Mediates Cell Corpse

RT Engulfment in C. elegans.";

RL Cell 104:43-56(2001).

DR EMBL: AL032657; CAA21739.1;

DR EMBL: AF332568; AAG60061.1;

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR002174; Furin-like.



---